

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
~~please do not report the images to the~~  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 ; Search time 15 Seconds  
(without alignments)  
936.314 Million cell updates/sec

Title: US-09-839-136-2  
Perfect score: 3096  
Sequence: 1 MRPWGSMRWIMLILFAMGT.....YKYREKIETVKVPTYPEADK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTOUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfillseq1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3088	99.7	575	US-08-913-805A-2	Sequence 2, Appl1
2	3088	99.7	575	US-09-442-629-2	Sequence 2, Appl1
3	2949	95.3	575	US-08-913-805A-10	Sequence 10, Appl1
4	2949	95.3	575	US-09-442-629-10	Sequence 10, Appl1
5	131	4.2	26	US-08-913-805A-3	Sequence 3, Appl1
6	131	4.2	26	US-09-442-629-3	Sequence 3, Appl1
7	126	4.1	25	US-08-913-805A-12	Sequence 12, Appl1
8	126	4.1	25	US-09-442-629-12	Sequence 12, Appl1
9	112.5	3.6	1036	US-08-968-752B-4	Sequence 4, Appl1
10	112	3.6	352	US-08-630-915A-12	Sequence 12, Appl1
11	111	3.6	1588	PCT-US93-07261-11	Sequence 11, Appl1
12	111	3.6	1663	PCT-US93-07261-16	Sequence 16, Appl1
13	110.5	3.6	971	US-09-107-149-19	Sequence 19, Appl1
14	110.5	3.6	1074	US-08-768-147B-2	Sequence 2, Appl1
15	110.5	3.6	1074	US-08-968-752B-2	Sequence 2, Appl1
16	110.5	3.6	1074	US-09-107-149-3	Sequence 3, Appl1
17	110	3.6	976	US-09-104-324B-4	Sequence 4, Appl1
18	110	3.6	3248	US-08-353-700-1	Sequence 1, Appl1
19	110	3.6	3248	PCT-US93-16216-1	Sequence 1, Appl1
20	109	3.5	214	PCT-US96-08950-2	Sequence 2, Appl1
21	109	3.5	229	PCT-US96-09127-2	Sequence 2, Appl1
22	103.5	3.3	677	US-08-630-915A-221	Sequence 221, App
23	103.5	3.3	1197	US-08-836-567-2	Sequence 12, Appl1
24	103	3.3	1230	US-08-968-542C-35	Sequence 35, Appl1
25	102	3.3	251	US-08-630-915A-8	Sequence 8, Appl1
26	98	3.2	247	US-08-937-271-13	Sequence 13, Appl1

28	98	3.2	343	US-08-937-271-17
29	97.5	3.1	676	US-09-313-930-2
30	97	3.1	2482	US-08-328-254-6
31	97	3.1	2616	5206163-3
32	96.5	3.1	546	US-08-630-915A-14
33	96.5	3.1	1052	US-08-863-118-2
34	96	3.1	1566	US-08-687-956A-23
35	95.5	3.1	546	US-08-942-423-4
36	95.5	3.1	1052	US-08-863-118-1
37	95.5	3.1	1052	US-09-377-310-2
38	95	3.1	922	US-09-141-206-6
39	94.5	3.1	1053	US-08-863-118-3
40	94.5	3.1	1375	US-09-722-139-2
41	94	3.0	933	US-09-141-206-2
42	94	3.0	933	US-09-107-149-17
43	94	3.0	2544	US-08-576-626A-32
44	93	3.0	933	US-09-107-149-2
45	92	3.0	344	US-09-298-886-11

## ALIGNMENTS

RESULT 1  
US-08-913-805A-2  
: Sequence 2, Application US/08913805A  
: Patent No. 6054304  
: GENERAL INFORMATION:  
: APPLICANT: TANIGUCHI, Naoyuki  
: APPLICANT: UOZUMI, Naofumi  
: APPLICANT: SHIBA, Tetsuo  
: APPLICANT: YAMAGUCHI, Shusaku  
: TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
: NUMBER OF SEQUENCES: 13  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kenyon & Kenyon  
: STREET: 1025 Connecticut Avenue, N.W., Suite 600  
: CITY: Washington  
: STATE: DC  
: COUNTRY: US  
: ZIP: 20036  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3+ Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
: SOFTWARE: WordPerfect 6.1 Windows  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/913, 805A  
: FILING DATE: 7 JAN 1998  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/JP97/00171  
: FILING DATE: 23 JAN 1997  
: APPLICATION NUMBER: JP 192260  
: FILING DATE: 22 JUL 1996  
: APPLICATION NUMBER: JP 162813  
: FILING DATE: 24 JUN 1996  
: APPLICATION NUMBER: JP 161648  
: FILING DATE: 21 JUN 1996  
: APPLICATION NUMBER: JP 10365  
: FILING DATE: 24 JAN 1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Toffenetti, Judith L.  
: REGISTRATION NUMBER: 35,048  
: REFERENCE/DOCKET NUMBER: 2356/3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 202-429-1776  
: TELEFAX: 202-429-0796  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 575 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear

Sequence 17, Appl  
Sequence 2, Appl1  
Sequence 6, Appl1  
Patent No. 5206163  
Sequence 14, Appl  
Sequence 2, Appl1  
Sequence 21, Appl  
Sequence 4, Appl1  
Sequence 1, Appl1  
Sequence 2, Appl1  
Sequence 6, Appl1  
Sequence 3, Appl1  
Sequence 2, Appl1  
Sequence 17, Appl  
Sequence 35, Appl  
Sequence 2, Appl1  
Sequence 11, Appl





RESULT 3  
US-08-913-805A-10  
Sequence 10, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-0796  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-10

Query Match 95.3% Score 2949; DB 3: Length 575;  
Best Local Similarity 95.5% Pred. No. 1.1e-276;  
Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPTGSRWMLILFAGTLLFYIGCHLVNDHSDHSSRELKTLAKLERLKOQNDL 60  
DB 1 MRPTGSRWMLILFAGTLLFYIGCHLVNDHSDHSSRELKTLAKLERLKOQNDL 60  
QY 61 RRMAGSLIPGPIDOGASGRVRALEQFMKAKQIENYKQTKNGCGKHILRRRIE 120  
DB 61 RRMAGSLIPGPIDOGASGRVRALEQFMKAKQIENYKQTKNGCGKHILRRRIE 120  
QY 121 NGAKELWFFLOSELKKLNLGSENELORHADFLSDLGHHRSIMTDLYLSOTDAGGMR 180  
DB 121 NGAKELWFFLOSELKKLNLGSENELORHADFLSDLGHHRSIMTDLYLSOTDAGGMR 180  
QY 181 EKAANDLTELVRITLYLQNPDCSKAKKLVGNINKGCGYCOLHHVVYCFMIAVGTORT 240  
DB 181 EKAANDLTELVRITLYLQNPDCSKAKKLVGNINKGCGYCOLHHVVYCFMIAVGTORT 240

DB 181 EKAANDLTELVRITLYLQNPDCSKAKKLVGNINKGCGYCOLHHVVYCFMIAVGTORT 240  
QY 241 LALESHNMRYATGMEVFRVSECTCDRSSSTGIMHSGEYKDKNNVOVELPIYDSVHPR 300  
DB 241 LILESQNMRYATGMEVFRVSECTCDRSSSTGIMHSGEYKDKNNVOVELPIYDSVHPR 300  
QY 301 PPLYPLAVPEDLADRLVAVHGDPAVMWVSQFVKYLIRPQPLKEKEIEBATEKLGFKHPVI 360  
DB 301 PPLYPLAVPEDLADRLVAVHGDPAVMWVSQFVKYLIRPQPLKEKEIEBATEKLGFKHPVI 360  
QY 361 GVHVRFTDKVGAFAAHPHIEEYTVHVEEDFQLLARMOVDKRRVYLATDDBALLKEAKTK 420  
DB 361 GVHVRFTDKVGAFAAHPHIEEYTVHVEEDFQLLARMOVDKRRVYLATDDBALLKEAKTK 420  
QY 421 YPSTEFISDNTSWSAGIHNRTENSLRGVLLDHLFSLQADFLVCTESSQYCRATEIMQ 480  
DB 421 YPSTEFISDNTSWSAGIHNRTENSLRGVLLDHLFSLQADFLVCTESSQYCRATEIMQ 480  
QY 481 ALHPDASANFRSLDDIYFEGPNAHNOIAIYPHOPRTGELPMEPGDIIAGVGNHMDGYP 540  
DB 481 TLHPDASANFRSLDDIYFEGPNAHNOIAIYAHOPRTADELPMEPGDIIAGVGNHMDGYP 540  
QY 541 KGVNRKLGRTGLYPSYKVRKIEYTVKYPPTYEADK 575  
DB 541 KGVNRKLGRTGLYPSYKVRKIEYTVKYPPTYEADK 575

RESULT 4  
US-09-442-629-10  
Sequence 10, Application US/09442629  
Patent No. 6291219  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NO. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-442-629-10

Query Match 95.3%; Score 2949; DB 4; Length 575;  
Best Local Similarity 95.5%; Pred. No. 1.1e-276;  
Matches 349; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPWTSWIMILFLPAMGTLFLYIGSHLVKRDNDSDHSSRELSTKLAKLERLKQONEDL 60  
1 MRPWTSWIMILFLPAMGTLFLYIGSHLVKRDNDSDHSSRELSTKLAKLERLKQONEDL 60  
DB 1 MRPWTSWIMILFLPAMGTLFLYIGSHLVKRDNDSDHSSRELSTKLAKLERLKQONEDL 60  
QY 61 RRMASGLRIPREGIDQSPASGRVRALEBPMAKKEQIENYKQTKNGPKGDHEILRRRIE 120  
61 RRMASGLRIPREGIDQSPASGRVRALEBPMAKKEQIENYKQTKNGPKGDHEILRRRIE 120  
DB 61 RRMASGLRIPREGIDQSPASGRVRALEBPMAKKEQIENYKQTKNGPKGDHEILRRRIE 120  
QY 121 NGAKELMFLQSELSKKLNKNEGNELOKHADEFLSDGHHHRSIMTDLYLSTQDGAQDMR 180  
121 NGAKELMFLQSELSKKLNKNEGNELOKHADEFLSDGHHHRSIMTDLYLSTQDGAQDMR 180  
DB 121 NGAKELMFLQSELSKKLNKNEGNELOKHADEFLSDGHHHRSIMTDLYLSTQDGAQDMR 180  
QY 181 EKEAKDLTELQRRITLYLQNPKDCSKAKKLVCNINKGCGCOLHHVVCFMATAGTORT 240  
181 EKEAKDLTELQRRITLYLQNPKDCSKAKKLVCNINKGCGCOLHHVVCFMATAGTORT 240  
DB 181 EKEAKDLTELQRRITLYLQNPKDCSKAKKLVCNINKGCGCOLHHVVCFMATAGTORT 240  
QY 241 LALESHNRKATGQWETVFRPVSETCTDRSGSTGHWGCVKKNQVVELPYDVSVHPR 300  
241 LALESHNRKATGQWETVFRPVSETCTDRSGSTGHWGCVKKNQVVELPYDVSVHPR 300  
DB 241 LALESHNRKATGQWETVFRPVSETCTDRSGSTGHWGCVKKNQVVELPYDVSVHPR 300  
QY 301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEEATKKGFEKHPVI 360  
301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEEATKKGFEKHPVI 360  
DB 301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEEATKKGFEKHPVI 360  
QY 361 GVHVRRTDKVGAFAHPRIEYTVHVEEDPQLLARRMOYDKKRVYLATDPPALLKEATK 420  
361 GVHVRRTDKVGAFAHPRIEYTVHVEEDPQLLARRMOYDKKRVYLATDPPALLKEATK 420  
DB 361 GVHVRRTDKVGAFAHPRIEYTVHVEEDPQLLARRMOYDKKRVYLATDPPALLKEATK 420  
QY 421 YPSEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCFSSQVCVAYEIMQ 480  
421 YPSEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCFSSQVCVAYEIMQ 480  
DB 421 YPSEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCFSSQVCVAYEIMQ 480  
QY 481 ALHPDASANFRSDDIYFEGCPNAHNOIATYPPHQPRTGEPHPEPDIISVAGNHMDGYP 540  
481 ALHPDASANFRSDDIYFEGCPNAHNOIATYPPHQPRTGEPHPEPDIISVAGNHMDGYP 540  
DB 481 ALHPDASANFRSDDIYFEGCPNAHNOIATYPPHQPRTGEPHPEPDIISVAGNHMDGYP 540  
QY 541 KGVNRKLGRTGLPSYKVRKEITVRYPTYPEADK 575  
541 KGVNRKLGRTGLPSYKVRKEITVRYPTYPEADK 575  
DB 541 KGVNRKLGRTGLPSYKVRKEITVRYPTYPEADK 575

RESULT 5  
US-08-913-805A-3  
Sequence 3, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naotumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-3

Query Match 4.2%; Score 131; DB 3; Length 26;  
Best Local Similarity 100.0%; Pred. No. 4e-06;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 KQTKNGPKGDHEILRRRIENGAKEL 126  
1 KQTKNGPKGDHEILRRRIENGAKEL 126  
DB 1 KQTKNGPKGDHEILRRRIENGAKEL 126  
RESULT 6  
US-09-442-629-3  
Sequence 3, Application US/09442629  
Patent No. 6291219  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naotumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NO. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-442-629-3

Query Match 4.2%; Score 131; DB 4; Length 26;  
Best Local Similarity 100.0%; Pred. NO. 4e-06;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 KOTKNGPKDHEILRRRIENGAKEL 126  
DB 1 KOTKNGPKDHEILRRRIENGAKEL 25  
|||||

RESULT 7  
US-08-913-805A-12  
Sequence 12, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913.805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-12

Query Match 4.1%; Score 126; DB 3; Length 25;  
Best Local Similarity 96.0%; Pred. NO. 1.2e-05;  
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 352 KLGFKHPVIGVHVRTDKVGAEAF 376  
DB 1 KLGFKHPVIGVHVRTDKVGTEAF 25  
|||||

RESULT 8  
US-09-442-629-12  
Sequence 12, Application US/09442629  
Patent No. 6291219  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442.629  
FILING DATE: 18-NO. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913.805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid



Query Match Similarity 3.6%; Score 112; DB 4; Length 352;  
Best Local Similarity 19.2%; Pred.No.0.017;  
Matches 81; Conservative 60; Mismatches 124; Indels 156; Gaps 18;

OY 178 DWREKAK-DLT-----ELVGRIRITYLQNFPCDSKAKLYCNI-----NKGCGYGCOL 224  
|::||::||::||::||::||::||::||::||::||::||  
Db 32 DKEMERKKVDYTSRAVMELMTKTIEYLQ-PNPASRAKLMSINTMSKITRGCKGPGYR-QA 89  
OY 225 HHVVVCSEMIAYGSTORTLALESHNMWRYATGWETVFRPVSETCTDRSGSSSTGHWSGEYKDK 284  
::||::||::||::||::||::||::||::||::||::||  
Db 90 EALLAEAMLKFG-----RELGDGC-----NFGPALGEVGEA 120  
OY 265 NVOVVELPLVDSDHNRPRPYLPALVPEDLDRLYRVNHODPRAWMVVSQFYKYTLIRQPWLK 344  
::||::||::||::||::||::||::||::||::||::||  
Db 121 MRLESE--VKDS-----LDMEVKONFTDPLNLHMKDL-----R 152  
OY 345 ELEEATKKLGKFHVIGVNHVRFDKGAFAHPRIEEY-----TVHVED 389  
||::||::||::||::||::||::||::||::||::||  
Db 153 ELQHHLKKEGRLDDGYKKKKRGGKIPDELRQALEKFDESKEIAESSMFLLEMDIEOV 212  
OY 390 FOLLA---RRMOYDKRKYVLATDPALAKE-----AKTKYP-----SYEFTSDNSIS 433  
||::||::||::||::||::||::||::||::||::||  
Db 213 SOLSALVAQOLEXHKQAVOILQGVTVLERIRIQASSQPREYQPKRMSLFPATGDTQ 272  
OY 434 WSAGLNRTENTSRLKVIIDIHFLSQADFLVCTFSSQVCRAVEITMALHPDASANPSL 493  
::||::||::||::||::||::||::||::||::||  
Db 273 PNGGLSHSTQPKR-PSVOMD-----OPCRALYDL----- 301  
OY 494 DDIYVFEGPNANHQLIIVPHOPTKEGPMPGPDIIIGVAGNHMGDPKGVARKLQRTGLY 553  
::||::||::||::||::||::||::||::||::||  
Db 302 -----EPENEGELAFKEGDITILTNOIDENWTYEGMLH--GQSGCF 339  
OY 554 P 554  
Db 340 P 340

RESULT 11  
PCT-US93-07261-11  
Sequence 11, Application PC/TUS9307261  
GENERAL INFORMATION:  
TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. C. Blasdale  
STREET: One Glralda Farms  
CITY: Madison  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07940-1000  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.5  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07261  
FILING DATE: 19930805  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/927,531  
FILING DATE: 07-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Blasdale, John H. C.  
REGISTRATION NUMBER: 31,895  
REFERENCE/DOCKET NUMBER: DX0288K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-822-7398  
TELEFAX: 201-822-7039  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1588 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

```

: MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match          3.6%; Score 111; DB 5; Length 1588;
Best Local Similarity 20.7%; Pred. No. 0.24;
Matches 103; Conservative 66; Mismatches 152; Indels 176; Gaps 24;

OY 24 YIGHLVNDNDSDSHSELKILAKLERLK-QQNDLRNAGSLRIPEDIDGPAASGR 82
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 FIKQHL---KDYEEKERKRRNWLRSRLRDLKRETEOLEKLNLALESAINELKERRASRR 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 83 VRALEEGPMKKKEQLEENKKO-----TKNPGCDHETLRRRIENGAKELMFLQSELRK 156
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 PMWVMQ-RGKMDVEDENIKKYDDEQAERKNGT-KDEELDKD--GDGYEEI--VETKFGY 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 137 LKNLEGNELQRHADFLSDLGHERSIMTDLYYLSQTDGAGDW-----179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 MR-----ENALGRLDEERYEKKRYLYKE-DGEDLDKVEKLEETGCFREK 288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 180 -----REKEAKDI/TELVRIRITYLQNPDKCSKAK-----KLVCINKGCGYGC 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 PFTTILVRKRRNKKQKTKLEDKEKKLLAABEDDEKKIKTLKSDDKVVPVNNK-----343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 223 QLHHVYVCPMALVGTQRTLALESNNKRY-ATGCGVTVRPVSEFCTDNRSGSGTGHMSGEV 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 -----KSSPDKFRAPDKKRTMFRLSBL-----367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 282 KDKNQVELPIVDGVHRRPPLPLAVAPEDLADRLVRYHG-----DP 323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 368 -----PRLV-----PRKDN-ELAVCGDSMDS--KNGKRLKLTNPFRKRRNKLKER 411
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 324 AVMWVSQVFKYLIRPOPMLEKEIEEATKKLGFKNHVGIVGVR-----TDKVGAEALHP- 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 412 KMQELHKFKKNKKYQKLLEREKRENPDEPLMPEI--HVIKPSDLMDKGNKSAGHPF 469
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 379 -----IEEY-TVHVEEDFOL-----LARKNQDKKRYLATDDPALLK 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 470 KYQPTKGLKEYESHVSVDYOLEHEPPKLPDEYKGVHSREYQDHH-----EPPTKLP 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 416 EAKTKYPSYEFTSDNSI 432
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 523 EYKGVHSREYQDNEV 539
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
PCT-US93-07261-16
: Sequence 16, Application PC/TUS9307261
: GENERAL INFORMATION:
: TITLE OF INVENTION: PLEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREO
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: John H. C. Blasdale
: STREET: One Giralda Farms
: CITY: Madison
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07940-1000
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 6.0.5
: SOFTWARE: Microsoft word 5.1a
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07261
: FILING DATE: 19930805
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/927,531
: FILING DATE: 07-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Blasdale, John H. C.
: REGISTRATION NUMBER: 31,895
: REFERENCE/DOCKET NUMBER: DX0288K
: TELECOMMUNICATION INFORMATION:
:

```

TELEPHONE: 201-822-7398  
TELEFAX: 201-822-7039  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1663 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: Malayan Camp  
PCT-US93-07261-16

Query Match	3.68;	Score 111;	DB 5;	Length 1663;
Best Local Similarity	20.7%;	Pred. No. 0.26;		
Matches 103;	Conservative 66;	Mismatches 152;	Indels 176;	Gaps 24

```

QY 24 YIGGHLVADNDHSDHSSSELSKILAKLERLK-QONEDLARMAGSLRIPREGPIODGSGSR 82
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 131 FIKQHL--KQYEEKKEKRRNMWILRSRLROKLRITBQLLEKLNQLESAINELERRASRR 187
QY 83 VRALBEDPMKKEQDIENTYKKO-----TKGPGCKDHEILRRRLNCGAKELMFLQSLKK 136
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 188 PMWYKMQ-RGKMKDEDEKMKYDDQEAQKNGT-KDEBIKDK--GDGYEII--VETKFGY 240
QY 137 LKNLEGNELQHNADFLSDGHNERSIMTDLUYLSOTDGAQDW----- 179
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 241 MR-----ENALGEDEYEEEREKKRYLKE-DGEGDCLKDVEBEKLEETGYGFRK 288
QY 180 -----REKEANDTELRYORITLYLQNPDKCSKAR-----KLVCNINKGCGYCC 222
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 FPTTRILVKKRRNKEQKKLKDEKKEKLLIAEEPDDEKKIKLKSDDDKVYVVPVKN----- 343
QY 223 QLNHVVCYCFMAYGTQRTLALSESINMYY-ATGCGMETVFRVYSECTDRSSSGHMSGEV 281
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 344 -----KSSPPDKFRAADKRTMFFYRLSEL----- 367
QY 282 KDKNVQVLEPIYDSVHPRPYRLPLAVPEDLADRLRVHNG-----DP 333
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 368 -----PPIV-----PRKN--ELAVCGDSMDS--KVNKKLKTSTPNPKRRRNKKLER 411
QY 324 AVMWVSQPVKLLIPROPWLKEIEEATKKLGFKHRYGVAVRR-----TDKVGAEAFHP- 378
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 412 KMOELHKKKKKKYKKYOKLLEKREKRENDPEPLNTPEI--HVIRSDMLMDGKNSGHP 469
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 379 -----IEEY-THVEEDFOL-----LARRMOYDKKKRYLATDDPALK 415
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 470 KYQPTKGLKEVEESHVSNDQYLDHEPPTKLPEYEEKGHVSREYOLDH-----EPPTKL 522
QY 416 EAKTKYPSYEFISDNT 432
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 523 EYKGVHSREYOLDNEV 539

RESULT 13
US-09-107-149-19
: Sequence 19, Application US/09107149
: Patent No. 6274363
: GENERAL INFORMATION:
: APPLICANT: Leung, David W.
: APPLICANT: Tompkins, Christopher K.
: TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
: FILE REFERENCE: 077319/0144
: CURRENT APPLICATION NUMBER: US/09/107,149
: CURRENT FILING DATE: 1998-06-30
: EARLIER APPLICATION NUMBER: 08/768,147
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 971
: TYPE: PRT
: ORGANISM: Homo sapiens

```

US-09-107-149-19	
Query Match	3.6%; Score 110.5; DB 4; length 971;
Best Local Similarity	18.0%; Pred.No. 0.12;
Matches 101; Conservative	67; Mismatches 166; Indels 227; Gaps 24.

```

OY 114 ILRRRIENGAR-----ELWFELOSLEKK--LKNLEGN-ELQIRADEFLSDL---GH 158
Db 404 ILKKAQOGVAFILFMILMYKEVELALGINSEYTKRFLMLRHPRIKMKRIPRDNVSTVYLIMAH 453
OY 159 HERSIMTD--LVLYSQTDGA--GDWREKAKDTELVOIRITTYLONPRDCSARKKLYCNIIN 215
Db 464 HEKLIYIDQSAVAFVGIDILAVGRMDNE-----HRLTDVGSVKRYTSGPSL----- 509
OY 216 KGGCGGOLHHVVCYFMALAYGTQRTLALESNNMRYATGCMGTVPYRRPSEICTBRSGSSTG 275
Db 510 -----GSLPPAAMESME-----521
OY 276 HWSGVEKKKNVQVELPI-----VDS-----296
Db 522 --SLRLKKNRPYONLPIQKSIDVDYDSKLGKIGKPRKFSKSLYKOLHRIHLIDADSISS 579
OY 297 -----VHRRPRLPLAVPEDLADR-LVR-----VHG 321
Db 580 IDSTSSYFNHYRSHHNLHGLKPHFKLFHPSESEOGILTRPHADTGSIRSLQTVGVELHG 639
OY 322 DPAWVWWSQFVKYLLRPPRYLEKELEE-----ARKKGFKIRPVGVNVRTDKVAGAEAR 376
Db 640 ETRFMHGKDYCNFNFVKDWQLDKPFPADEIDRYSTPRMWMHDIASAVI---GRAADVAR 695
OY 377 HPTEYTVHVEEDPOLLARROQVKKRYVLYATDTPALALKEAKTKYPSYFEISDNI----432
Db 696 HFIOQM-----NFTKI-----MKSYRSLSTPRLPSPQTLAHBLRYQVGSVHANV 742
OY 433 -----SWSAGLHNRYTENSLEGVYIDI-----HFLSQADFLVCTESSQVC 472
Db 743 QLLRSAAWMSAGI--KYHEES-IAAUYVIENSRRHYIENQFISCADKV--VFNKIGD 799
OY 473 RYAVYIMDA-----LHPDASANFRSLDIDYFGGNNAINOIAIYRPHOPRTE 518
Db 800 ALAQIILAHHEENOKRYRYVYIPLLPFEGEDIST-----GGNALQAIIMHFNRYKIMCR 852
OY 519 GEIPMEPGDIIGVAGNHWDGY 539
Db 853 GENSI-LGQKAELGNQWINY 872

RESULT 14
US-08-768-147B-2
: Sequence 2, Application US/08768147B
: Patent No. 5859222
: GENERAL INFORMATION:
: APPLICANT: Leung, David W.
: TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cell Therapeutics, Inc.
: STREET: 200 Elliott Avenue West, Suite 400
: CITY: Seattle
: STATE: Washington
: COUNTRY: U.S.A.
: ZIP: 98119
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" disk, 1.44mb, double side, high density
: OPERATING SYSTEM: Hewlett Packard-IBM Compatible
: SOFTWARE: WORD for WINDOWS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/768,147B
: FILING DATE: 16-Dec-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION: 435

```

```

: APPLICATION NUMBER: Provisional US6008768
: FILING DATE: 15-Dec-1995
: ATTORNEY/AGENT INFORMATION:
:   NAME: Paciszewski, Stephen
:   REGISTRATION NUMBER: 36,131,
:   REFERENCE/DOCKET NUMBER: 1802A
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: (206)282-7100
:     TELEFAX: (206)284-6206
: INFORMATION FOR SEQ ID NO: 2:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 1074
:     TYPE: amino acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:   MOLECULE TYPE: polypeptide
:   HYPOTHETICAL: no
:   ANTI-SENSE: no
:   FRAGMENT TYPE:
:   ORIGINAL SOURCE:
:     ORGANISM: homo sapien
:     STRAIN:
:     INDIVIDUAL ISOLATE:
:     DEVELOPMENTAL STAGE:
:     HAPLOTYPE:
:     TISSUE TYPE:
:     CELL TYPE:
:     ORGANELLE:
:   IMMEDIATE SOURCE:
:     LIBRARY:
:     CLONE:
:   POSITION IN GENOME:
:     CHROMOSOME/SEGMENT:
:     MAP POSITION:
:   UNITS:
:   FEATURE:
:     NAME/KEY: hPLD1
:     LOCATION:
:     IDENTIFICATION METHOD:
:     OTHER INFORMATION:
:   PUBLICATION INFORMATION:
:     AUTHORS:
:     TITLE:
:     JOURNAL:
:     VOLUME:
:     ISSUE:
:     PAGES:
:     DATE:
:     DOCUMENT NUMBER:
:     FILING DATE:
:     PUBLICATION DATE:
:     RELEVANT RESIDUES IN SEQ ID NO: 2:
: US-08-768-147B-2

```

[illegible]

```

Db 522 ---SLRLKDNPEVONPLRLQKSIDDDVDSKLKGTGPKRKFSSKSLYKQLRHNLHLDADSISS 579
QY 297 -----VHRRPPLPLAVPEDIADR-LVR-----VHG 321
Db 580 IDSTSSYFNHYRSHNHLIHGLKPHFKLPLTPRSEBEOGLTRRHADGSIKSLQGVGLHG 639
QY 322 DPAVWWWQGFUKYLLRLDRPYLEKETEE-----ATKLGFKHPRVIGVHVRRTDKVGAELAF 376
Db 640 ETRFMHGKDXCNPFYFKDVLQDJKPFAFDIDRYSTRPRMHDIASAVH-----GKARQVAR 695
QY 377 HPIEETVHVCEDQQLLARHQVQDKKRVUYLATDDRPALLKAKTTPSYEFISOMSI----- 432
Db 696 HFIOW-----NFKTI-----MKSRYRSLSTRPLPSOTANHELRLQVPSVANY 742
QY 433 -----SMSAGLHNRYTENSLRGVLIIDI-----HFLSADPLVCTFSSQVC 472
Db 743 QILRSADMSWAGCI--KHNEBSIHAAVYHVIENSRHYIYIENQFISQADKDV-VFNKIGD 799
QY 473 RVAYEIMQA-----LHPDASANFSLDDIYFFGGFNAHQALAYPHORPTE 518
Db 800 AIAORILKAHREKOKRYRVYVILPRLPGEGEDIST-----GGNALQDAIWHFNYPTMCR 852
QY 519 GEIPMEPEDIIGVAGNHWDGY 539
Db 853 GENSI-IGQLKAEIAGNQMINT 872

```

RESULT 15  
 US-08-968-752B-2  
 Sequence 2, Application US/08968752B  
 Patent No. 6043073  
 GENERAL INFORMATION:  
 APPLICANT: Frohman, Michael A.  
 APPLICANT: Morris, Andrew  
 TITLE OF INVENTION: No. 6043073el Phospholipase D Polypeptide and  
 TITLE OF INVENTION: DNA Sequences  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: ONYX Pharmaceuticals, Inc.  
 STREET: 3031 Research Drive  
 CITY: Richmond  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94806  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/968,752B  
 FILING DATE: 13-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/025,469  
 FILING DATE: 05-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Giotta, Gregory J.  
 REGISTRATION NUMBER: 32,028  
 REFERENCE/DOCKET NUMBER: ONYX2004  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 510-222-9700  
 TELEFAX: 510-222-9758  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1074 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-968-752B-2





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:00 ; Search time 35 Seconds  
(without alignments)  
1824.784 Million cell updates/sec

Title: US-09-839-136-2

Perfect score: 3096

Sequence: 1 MRWGTGSMRWIMILIFAMGT.....YKVRKIEYKVPYTPPEADK 575

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3088	99.7	575	18	AAW22124
2	2949	95.3	575	18	AAW22125
3	2610	84.3	515	22	AAW22125
4	1732	55.9	339	22	AAW22125
5	1431.5	46.2	223	22	AAW22125
6	1168	37.7	233	22	AAW22125
7	438	14.1	82	22	AAW22125
8	438	14.1	82	22	AAW22125
9	438	14.1	82	22	AAW22125
10	438	14.1	82	22	AAW22125
11	438	14.1	82	22	AAW22125

12	438	14.1	82	22	AAW15003
13	438	14.1	82	22	AAW27440
14	438	14.1	82	22	AAW2730
15	432	14.0	81	22	AAW4859
16	432	14.0	81	22	AAW37789
17	131	4.2	26	18	AAW22126
18	126	4.1	25	18	AAW22131
19	123	4.0	1951	22	AAW78835
20	120.5	3.9	1427	22	AAW10534
21	119	3.8	1456	22	AAW58673
22	119	3.8	1711	22	AAW79819
23	118.5	3.8	1392	20	AAW06999
24	117.5	3.8	2400	22	ABG20278
25	116.5	3.8	2415	22	ABG20279
26	113.5	3.7	1097	22	ABG57814
27	112.5	3.6	1036	19	AAW53280
28	112	3.6	352	17	AAW05388
29	112	3.6	1527	22	AAW01184
30	111	3.6	398	21	AAW31272
31	111	3.6	437	21	AAW6252
32	111	3.6	1588	15	AAW46605
33	111	3.6	1663	15	AAW46608
34	110.5	3.6	971	22	AAW44717
35	110.5	3.6	971	22	AAW85981
36	110.5	3.6	1074	19	AAW53279
37	110.5	3.6	1074	22	AAW85976
38	110.5	3.6	1125	22	ABG18927
39	110	3.6	976	22	AAW6581
40	110	3.6	981	22	AAW36618
41	110	3.6	3248	17	AAW9795
42	109.5	3.5	762	22	AAW57227
43	109	3.5	214	18	AAW09037
44	109	3.5	214	18	AAW12706
45	109	3.5	229	17	AAW05403

#### ALIGNMENTS

RESULT 1	
AAW22124	
ID	AAW22124 standard; Protein: 575 AA.
XX	
AC	AAW22124;
XX	
DT	05-MAR-1998 (first entry)
XX	
DE	Pig alpha 1-6 fucosyltransferase.
XX	
KW	Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
KW	guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW	GlcNAc; cancer diagnosis.
XX	
OS	Sus scrofa.
XX	
PN	W09727303-A1.
XX	
PD	31-JUL-1997.
XX	
PE	23-JAN-1997; 97WO-JP00171.
XX	
PR	22-JUL-1996; 96JP-0192260.
PR	24-JAN-1996; 96JP-0010365.
PR	21-JUN-1996; 96JP-0161648.
PR	24-JUN-1996; 96JP-0162813.
XX	
PA	(TOYM ) TOYO BOSEKI KK.
XX	
PI	Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
XX	
DR	WPI, 1997-393690/36.
DR	N-PSDB; AAT76573.
XX	

Peptide #1437 enco  
Peptide #1477 enco  
Peptide #1412 enco  
Human brain expres  
Peptide #11826 enc  
Fragment #1 of pig  
Fragment #2 of hum  
Human protein SEQ  
Human 160KD mediat  
Drosophila melanog  
Human protein SEQ  
Resilin protein seq  
Novel human diagno  
Novel human diagno  
Drosophila melanog  
Human phospholipas  
Human phospholipas  
Mouse SH3P4-prote  
Rat glutamate tran  
Arabidopsis thalia  
Arabidopsis thalia  
Malaria P1EMP3 ep  
Plasmodium falcipa  
Human phosphatidyl  
Human phosphatidyl  
Human phospholipas  
Human phosphatidyl  
Novel human diagno  
Human SCP-1 muteln  
Staphylococcus aur  
Kinetochore prote  
Protonbacterium  
Osteoclast stimula  
Osteoclast stimula  
Human clone 5 prot

PT Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for  
PT synthesis and modification of sugar chains and used as an antigen  
PT for production of diagnostic antibodies  
PS Claim 4: Page 30-34; 61pp: Japanese.  
XX  
XX  
CC AAM22124 and AAM22125 represent the pig and human alpha 1-6  
CC fucosyltransferases of the invention, respectively. The enzyme transfers  
CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
CC nearest to R in the receptor molecule: (GlcNAc-beta 1-2Man-alpha 1-6)  
CC (GlcNAc-beta 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4GlcNAc-R to give  
CC 1-4GlcNAc-beta 1-4(Fuc-alpha 1-6)(GlcNAc-R. It has an optimum pH of about  
CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
CC hours at 4 degrees C. The optimum working temperature of the  
CC alpha 1-6 fucosyltransferase is 30-37 degrees C. A bivalent metal is  
CC not required for activity of the enzyme, and the enzyme is not inhibited  
CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
CC modification of sugar chains, and as antigen for the production of  
CC antibodies recognising the enzyme. The antibodies can be used for the  
CC diagnosis of cancer and other diseases.  
XX  
XX

SO Sequence 575 AA:

Query Match 99.7%; Score 3088; DB 18; Length 575;  
Best Local Similarity 99.8%; Pred. No. 5,7e-289;  
Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60  
DB 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60  
OY 61 RRMAGSLRIPEGIDPGPAGGRVRALEEQFMKAKEQIENYKOTKNGPKGDHILRRRIE 120  
DB 61 RRMAGSLRIPEGIDPGPAGGRVRALEEQFMKAKEQIENYKOTKNGPKGDHILRRRIE 120  
OY 121 NGAKELWFLQSELSKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180  
DB 121 NGAKELWFLQSELSKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180  
OY 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVYCNINNGCGGCOLHHVYCFMAYGTQRT 240  
DB 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVYCNINNGCGGCOLHHVYCFMAYGTQRT 240  
OY 241 LALESNHMRATGMEVFERPVSETCTDRSGSGTHMSGCVKKNQVVELPIYDSVHPR 300  
DB 241 LALESNHMRATGMEVFERPVSETCTDRSGSGTHMSGCVKKNQVVELPIYDSVHPR 300  
OY 301 PLYPLAIVEDLADRLVRVHGDPVAVMWVSQFVKYLLIRPOFWLEKEIEATKKLGFKHPVI 360  
DB 301 PLYPLAIVEDLADRLVRVHGDPVAVMWVSQFVKYLLIRPOFWLEKEIEATKKLGFKHPVI 360  
OY 361 GVAVRRPDKGAAAFPIEETVHVHEEDFOLLARMOVKKRVYLATDPAALKEAKTK 420  
DB 361 GVAVRRPDKGAAAFPIEETVHVHEEDFOLLARMOVKKRVYLATDPAALKEAKTK 420  
OY 421 YPSYEFTSDSISMSAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSOVCVAAVEIMQ 480  
DB 421 YPSYEFTSDSISMSAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSOVCVAAVEIMQ 480  
OY 481 ALHPDASANFRSLDDIYFEGCPAHHNOIATYHPQRTBEGTIPMEPGDITGACNHHMDGP 540  
DB 481 ALHPDASANFRSLDDIYFEGCPAHHNOIATYHPQRTBEGTIPMEPGDITGACNHHMDGP 540  
OY 541 KGVNRKLGRTGLPSYKVRKEITGVKPYTPPEADK 575  
DB 541 KGVNRKLGRTGLPSYKVRKEITGVKPYTPPEADK 575

RESULT 2  
AAM22125  
ID AAM22125 standard; Protein: 575 AA.  
XX

AC AAM22125;  
XX  
XX 05-MAR-1998 (first entry)  
DE Human alpha 1-6 fucosyltransferase.  
XX  
XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
KM GlcNAc; cancer diagnosis.  
OS Homo sapiens.  
PN MO9727303-A1.  
XX 31-JUL-1997.  
XX 23-JAN-1997; 97WO-JP00171.  
XX  
XX 22-JUL-1996; 96JP-0192260.  
PR 24-JAN-1996; 96JP-0010365.  
PR 21-JUN-1996; 96JP-0161648.  
PR 24-JUN-1996; 96JP-0162813.  
XX  
XX (TOYM ) TOYO BOSEKI KK.  
PA  
PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
XX  
XX MPI; 1997-393690/36.  
XX N-PSDB; AAT76574.  
XX  
XX Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for  
PT synthesis and modification of sugar chains and used as an antigen  
PT for production of diagnostic antibodies  
XX  
XX  
PS Claim 17: Page 39-43; 61pp: Japanese.

XX  
XX AAM22124 and AAM22125 represent the pig and human alpha 1-6  
XX fucosyltransferases of the invention, respectively. The enzyme transfers  
XX fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
XX nearest to R in the receptor molecule: (GlcNAc-beta 1-2Man-alpha 1-6)  
XX (GlcNAc-beta 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4GlcNAc-R to give  
XX 1-4GlcNAc-beta 1-4(Fuc-alpha 1-6)(GlcNAc-R. It has an optimum pH of about  
XX 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
XX hours at 4 degrees C. The optimum working temperature of the  
XX alpha 1-6 fucosyltransferase is 30-37 degrees C. A bivalent metal is  
XX not required for activity of the enzyme, and the enzyme is not inhibited  
XX in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
XX modification of sugar chains, and as antigen for the production of  
XX antibodies recognising the enzyme. The antibodies can be used for the  
XX diagnosis of cancer and other diseases.  
XX

SO Sequence 575 AA:

Query Match 95.3%; Score 2949; DB 18; Length 575;  
Best Local Similarity 95.5%; Pred. No. 1,5e-275;  
Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

OY 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60  
DB 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60  
OY 61 RRMAGSLRIPEGIDPGPAGGRVRALEEQFMKAKEQIENYKOTKNGPKGDHILRRRIE 120  
DB 61 RRMAGSLRIPEGIDPGPAGGRVRALEEQFMKAKEQIENYKOTKNGPKGDHILRRRIE 120  
OY 121 NGAKELWFLQSELSKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180  
DB 121 NGAKELWFLQSELSKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180  
OY 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVYCNINNGCGGCOLHHVYCFMAYGTQRT 240  
DB 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVYCNINNGCGGCOLHHVYCFMAYGTQRT 240

QY	241	LALESHNNRYATGAGNETVPEPVSECTDSSGSGTGHMGSQVKKANQVLEPLTVDSVHR	300
Db	241	LILESQNNRYATGAGNETVPEPVSECTDSSGSGTGHMGSQVKKANQVLEPLTVDSVHR	300
QY	301	PEYLPVAVPEDLADRLVRVHGDPVAVMWVSQFVKYLRLRPQWLEKEIEEATKKLGFKHPI	360
Db	301	PEYLPVAVPEDLADRLVRVHGDPVAVMWVSQFVKYLRLRPQWLEKEIEEATKKLGFKHPI	360
QY	361	GVNVRTDKVGEAAFNHPIEEYTVHNEEDPOLLARKMQDKKVVYLTATDPAALLKEAKTK	420
Db	361	GVNVRTDKVGEAAFNHPIEEYTVHNEEDPOLLARKMQDKKVVYLTATDPAALLKEAKTK	420
QY	421	YPSYEIISDNSTSMWSAGLNRYTENSLRQVYLLIDHFLSQADFLVCTFPSSQVCVAYEIMQ	480
Db	421	YPNYEFISDNSTSMWSAGLNRYTENSLRQVYLLIDHFLSQADFLVCTFPSSQVCVAYEIMQ	480
QY	481	ALHPDASANFSLDDIYFEGGPNANQOIAIYHPQPTREGIPEMPEGDILIGVAGNHMDGP	540
Db	481	TLHPDASANFSLDDIYFEGGPNANQOIAIYAHQPTRADEIPEMPEGDILIGVAGNHMDGS	540
QY	541	KGVNRRKLGRTGLYPSKYVREKIEFTVYKPYTPPEADK	575
Db	541	KGVNRRKLGRTGLYPSKYVREKIEFTVYKPYTPPEAK	575
RESULT 3			
ID	AAAG73884	standard; Protein; 515 AA.	
XX	AAAG73884;		
AC	AAAG73884;		
DT	03-SEP-2001	(first entry)	
XX			
DE	Human colon cancer antigen protein SEQ ID NO:4648.		
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; chromosome 14.		
XX			
OS	Homo sapiens.		
PN	WO200122920-A2.		
PD	05-APR-2001.		
XX			
PE	28-SEP-2000; 2000WO-US26524.		
XX			
PR	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
PA	(HUMA-). HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
DR	WPI; 2001-235357/24.		
DR	N-PSDB; AAH33315.		
XX			
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
PT	useful for preventing, diagnosing and/or treating colorectal cancers -		
PS	Claim 11; Page 6451-6453; 9803BP; English.		
XX			
CC	AAH32943 to AAH37195 and AA573514 to AA677788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patient's own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated Ps,		

Query Match	Best Local Similarity	Score	DB	Length
Matches 486; Conservative	95.3%;	2610;	7.1e-243;	515;
	6;	Mismatches 18;	Indels 0;	Gaps 0
66	SLRIIEGPIDGPGASGRVRLAEEOEFKAKEQIENYKKQTKNGPGKHDLRRRIENGAK	125		
6	SLRIIEGPIDGPGASGRVRLAEEOEFKAKEQIENYKKQTKNGPGKHDLRRRIENGAK	65		
126	LMFPIQSELKTKLKNLEGNLQHNADFLSDLGHNERSIMTDLYLLYSOTDAGGMRKEAK	185		
66	LMFPIQSELKTKLKNLEGNLQHNADFLSDLGHNERSIMTDLYLLYSOTDAGGMRKEAK	125		
186	DLTELVRITITLQNPDOCSKAKKLVLCNINKGCGCOLHNHYVCFMIAIGTORLALS	245		
126	DLTELVRITITLQNPDOCSKAKKLVLCNINKGCGCOLHNHYVCFMIAIGTORLALS	185		
246	HNMRVATGMEVTFRVPSECTDRSGSSYGHNSGVEYKDKNVQVELPIYDSVRRPRLP	305		
186	QNMVRATGMEVTFRVPSECTDRSGSSYGHNSGVEYKDKNVQVELPIYDSVRRPRLP	245		
306	LAVPEDLADRLVHNGDRAVVMVWSQFVKTLIRPQWLEKEIEBATKKGKFNPGVGVNR	365		
246	LAVPEDLADRLVHNGDRAVVMVWSQFVKTLIRPQWLEKEIEBATKKGKFNPGVGVNR	305		
366	RTDKVGAEEAFPIEEYTVHVEEDPOLLRNNOVKKRYVLTDDPALKLEAKTKPYSE	425		
306	RTDKVGAEEAFPIEEYTVHVEEDPOLLRNNOVKKRYVLTDDPALKLEAKTKPYSE	365		
426	FISDSISMSAGLHNRYTENSLSRGVILDIHFLSQADFLVCTFSQYCRVAYEIMQALHPD	485		
366	FISDSISMSAGLHNRYTENSLSRGVILDIHFLSQADFLVCTFSQYCRVAYEIMQALHPD	425		
486	ASANFRSLDDIYFGGPNHNOIATYPHQRTGELIPMERGDIIGVAGNHMDGYPGVNR	545		
426	ASANFRSLDDIYFGGPNHNOIATYPHQRTGELIPMERGDIIGVAGNHMDGYPGVNR	485		
546	KLGRTGLPSTYVREKIEYVKYPTYPEADK	575		
486	KLGRTGLPSTYVREKIEYVKYPTYPEADK	515		
Result 4	AAB75061			
ID	AAB75061	standard; Protein; 339 AA.		
XX	AAB75061;			
XX	20-JUL-2001	(first entry)		
XX	Human alpha 1-6 fucosyltransferase protein 237-575	SEQ ID NO:3.		
XX	Human: alpha 1-6 fucosyltransferase; alpha 1-6 fucN; antibody;			
XX	alpha 1-6 fucose transferase; anti-human alpha 1-6 fucN; immunossay.			
XX	Homo sapiens.			
XX	JP2001011097-A.			
XX	16-JAN-2001.			
XX	29-JUN-1999;	99JP-0183569.		

PR 29-JUN-1999; 99JP-0183569.  
XX  
PA (FJRE ) FUJIREBIO KK.  
XX  
DR WPI; 2001-275926/29.  
DR N-PSDB; AAE87952.  
XX  
PT Novel anti-human alpha1-6 fucose transferase antibody useful for  
XX immunosay -  
PS Example 2; Page 8-9; 11pp; Japanese.  
XX  
CC The present invention describes an anti-human alpha 1-6  
CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose  
CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunosay for detecting human  
CC alpha 1-6 fuct by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunosay. The anti-human  
CC alpha 1-6 fuct antibody can be used for immunosay. The present  
CC sequence represents the human alpha 1-6 fuct protein of residues 237  
CC to 575, which is used in an example from the present invention.  
XX  
SQ Sequence 339 AA:  
Query Match 55.9%; Score 1732; DB 22; Length 339;  
Best Local Similarity 95.0%; Pred. No. 2,3e-158;  
Matches 322; Conservative 4; Mismatches 13; Indels 0; Gaps 0;  
QY 237 TQRTLAESNNKRYAGCGTETFRPVSSETCTDSSGSTGMSGEVKDKNOVVELPIVDS 236  
DB 1 TORFLIESQNMRYATGCGTETFRPVSSETCTDSSGSTGMSGEVKDKNOVVELPIVDS 60  
QY 297 VHRRPPLPLAIVEDLADRLVRVHGDPVMMVSGFVYLLRPQPMLEKEIEATKRLGFK 356  
DB 61 LHRRPPLPLAIVEDLADRLVRVHGDPVMMVSGFVYLLRPQPMLEKEIEATKRLGFK 120  
QY 357 HPAVIGVHVRTDVGAAEAHPRIEETVYVHEEDPOLLARRMOYDKKRVYLATDDPALKE 416  
DB 121 HPAVIGVHVRTDVGAAEAHPRIEETVYVHEEHFOLLARRMOYDKKRVYLATDDPSLKE 180  
QY 417 AKRKYPEREISNSTSMSGLNHRYTENSLRGVYLDIHFLSQADPLVCFSSQVCHVAY 476  
DB 181 AKRKYPEREISNSTSMSGLNHRYTENSLRGVYLDIHFLSQADPLVCFSSQVCHVAY 240  
QY 477 EIMQALHPDASANFRSLDDIYFEGCPNANHQIAIYRPHQRTGEGEIPMEPGDIIGVAGNMH 536  
DB 241 EIMQALHPDASANFRSLDDIYFEGCPNANHQIAIYRPHQRTGEGEIPMEPGDIIGVAGNMH 300  
QY 537 DGYRPGVNRKLGRTGLYRSTYKVRKLETTVYKPYTPPEADK 575  
DB 301 DGYRPGVNRKLGRTGLYRSTYKVRKLETTVYKPYTPPEAEK 339  
RESULT 5  
ABB59117  
ID ABB59117 standard; Protein: 619 AA.  
XX  
AC ABB59117;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 4143.  
XX  
KM Drosophila: developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.

XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03220.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
PS Disclosure: SEQ ID NO 4143; 21pp + Sequence Listing: English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL161840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 619 AA:  
Query Match 46.2%; Score 1431.5; DB 22; Length 619;  
Best Local Similarity 44.4%; Pred. No. 6.4e-129;  
Matches 270; Conservative 110; Mismatches 179; Indels 49; Gaps 7;

QY 7 SW-RWMLIFANGTLLEFYTGHLVYRDN-----DHSDBSRRELKTLAKLERKQONEDL 60  
DB 14 SWARALIFVLAIGLVYFVVKLTNTOGOAAGESELNARISQALOMLEHTRQNEEL 73  
QY 61 RRMAGSLRIPEGRPIIDOGPASGRVAL-----EIOFMKAKQIOINRYK 102  
DB 74 KQIDELMSDQ--LDKOSAMKLVQRLNDALNPKLAEVAGRPRESMFESAPDLGWMNN 131  
QY 103 QTRNGPCK-----DHELTRRIENGAKELVFLQSELUKLKLNL---EG 142  
DB 132 VAGCAHPNDLAEVGRDGEFEPSLEYETRRRIQTNGICEINFEPSSEGLKRVAKVAGHNS 191  
QY 143 NELQRADEFLSDGHNHERSTIMTDLYLISOTDAGCMREKAKDLIELVQRTTYIQNRK 202  
DB 192 ADLEESTINOVLLQGADEHKRSLSDMERMRQSDGEAMRHRKEARDLSDLVQRRHLNHLQNP5 251  
QY 203 DCSKAKLVGNINKGCGYGCGQLHHVVYCFMAYGTQRTTLAESNNRYATGCGTETFRPVP 262  
DB 252 DCOMARKLVCKLNKGGCGYQGLHHVVYCFVAYATERTTLKRGKRYKKGMEVEYFORV 311  
QY 312 SENSCHDAGTANTYNNPCK---PMTQVLYLPIIDSLMPRPYPLAIVEDLAPRLKRLHSD 368  
QY 323 PAVMMVWSQFVKYLLRPQPMLEKEIEATKLGFKNHYVIGVHVRTDVGAAEAHPRIEY 382  
DB 369 PAVMMVWSQFVKYLLRPQPTTRDFLTSGMRNLGWERPIVGVHVRTDVGTEAACHSVEEY 428  
QY 383 TVHEEDFOLLARRMOYDKKRVYLATDDPALKEAKTKYPSYEFISDNTSMSAGLHNR 442  
DB 429 MIVYEDYKRTLEVNGSTVARIRFLASDDAOVIEARKKTYOYIIDPEYARAASTRY 486  
QY 443 TENSRLGVYLDIHFLSQADPLVCTFSSQVCRVAYEIMQALHPDASANFRSLDDIYFEGP 502  
DB 489 TDTALNGIYLDIHLLSMSDLVCTFSSQVCRVAYEIMQTMYPDAANFRSLDDIYFEGQ 548  
QY 503 NAHNQIAIYRPHQRTGEGEIPMEPGDIIGVAGNMWDGYPKGVNRKLGRTGLYRSTYKVRK 562  
LETTVYKPYTPPEADK 575

Db 549 NANNRRVIAHPRTHEDLGLRVGDLVSVAGNHMDGSKGNTRTNOGLFPSPFVEEKV 608  
QY 563 ETVKYPY 570  
: | | | |  
Db 609 DTAKLPY 616

RESULT 6  
AAB75062  
ID AAB75062 standard; Protein: 233 AA.

XX AAB75062;

XX 20-JUL-2001 (first entry)

XX Human alpha 1-6 fucosyltransferase protein 343-575 SEQ ID NO:4.

XX Human; alpha 1-6 fucosyltransferase; alpha 1-6 fuct; antibody;  
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fuct; immunoassay.

XX Homo sapiens.

XX JP2001011097-A.

XX 16-JAN-2001.

XX 29-JUN-1999; 99JP-0183569.

XX 29-JUN-1999; 99JP-0183569.

XX (FURE ) FUJIREBIO KK.

XX WPI: 2001-275926/29.

XX N-PSDB: AAF87953.

XX Novel anti-human alpha1-6 fucose transferase antibody useful for  
PT immunoassay -

XX Example 2: Page 9-10; 11pp: Japanese.

XX The present invention describes an anti-human alpha 1-6  
CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose  
CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fuct by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fuct antibody can be used for immunoassay. The present  
CC sequence represents the human alpha 1-6 fuct protein of residues 343  
CC to 575, which is used in an example from the present invention.

XX Sequence 233 AA:

Query Match 37.7%; Score 1168; DB 22; Length 233;  
Best Local Similarity 94.4%; Pred. No. 3.7e-104;  
Matches 220; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 343 EKEIEEATKKGFKHPYGVHVRRTDKVGAEEAFHPIEEYRVHVEDDQLARRMOVKK 402

Db 1 EKEIEEATKKGFKHPYGVHVRRTDKVGAEEAFHPIEEYRVHVEDDQLARRMOVKK 60

QY 403 RYVLTATDDPALKEAKTKYPSYEFISDNSISWSAGLHNRYTENSURGYLIDHFLSQADF 462

Db 61 RYVLTATDDPALKEAKTKYPSYEFISDNSISWSAGLHNRYTENSURGYLIDHFLSQADF 120

QY 463 LVCTFSSQVCRAVYEIFMALPQASANFRSLDDIYFEGCPNANHOIAIYHPQRTGEGIP 522

Db 121 LVCTFSSQVCRAVYEIFMALPQASANFRSLDDIYFEGCPNANHOIAIYHPQRTGEGIP 180

QY 523 MEPCGIIVAGNHMDGPKGVNRKLGRTGLYPSYKVRKEIEYKPYPEAK 575

Db 181 MEPCGIIVAGNHMDGPKGVNRKLGRTGLYPSYKVRKEIEYKPYPEAK 233

RESULT 7  
AAB28797  
ID AAB28797 standard; Peptide: 82 AA.

XX AAB28797;

XX 01-FEB-2002 (first entry)

XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 11765; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosis breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 82 AA:

Query Match 14.1%; Score 438; DB 22; Length 82;  
Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EYKDKNVQVELPIYDSVHPRPYLPLAVPEDLADRLVRHGDPAVMWSQFVKYLRPQ 339

Db 1 EYKDKNVQVELPIYDSVHPRPYLPLAVPEDLADRLVRHGDPAVMWSQFVKYLRPQ 60

QY 340 PWLEKEIEEATKKGFKHPYGV 361

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Db 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 8  
ID ABB33985 standard; Peptide; 82 AA.  
XX  
AC ABB33985;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #1491 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PS analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 26620; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 82 AA:  
XX  
XX Query Match 14.1%; Score 438; DB 22; Length 82;  
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EVKDKNVQVELPIYDSVHPRPYLPVAVPEDLADRLVHNGDPVWWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVELPIYDSVHPRPYLPVAVPEDLADRLVHNGDPVWWSQFVKYLIRPQ 60

QY 340 PWLEKEIEEATKKLGFKHPYIG 361  
DB 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 9  
ID ABB19422 standard; Protein; 82 AA.  
XX  
AC ABB19422;  
XX

DT 23-JAN-2002 (first entry)  
XX  
DE Protein #1421 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
PS Claim 15; SEQ ID NO 21192; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarray.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 82 AA:  
XX  
XX Query Match 14.1%; Score 438; DB 22; Length 82;  
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EVKDKNVQVELPIYDSVHPRPYLPVAVPEDLADRLVHNGDPVWWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVELPIYDSVHPRPYLPVAVPEDLADRLVHNGDPVWWSQFVKYLIRPQ 60

QY 340 PWLEKEIEEATKKLGFKHPYIG 361  
DB 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 10  
ID AAM54747 standard; Protein; 82 AA.  
XX  
AC AAM54747;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26852.  
XX

KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.  
XX Homo sapiens.  
XX MO200157275-A2.  
XX PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00667.  
XX PF 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX DR WPI; 2001-483446/52.  
XX XX  
XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX PT brains -  
XX PS Example 4; SEQ ID NO: 26352; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of  
XX CC the probes of the invention.  
XX SO Sequence 82 AA:  
XX  
XX Query Match 14.1%; Score 438; DB 22; Length 82;  
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 280 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPVAMVWSQFVKYLIRPQ 339  
XX DB 1 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPVAMVWSQFVKYLIRPQ 60  
XX QY 340 PWLEKEIEATKKGKHPVIG 361  
XX DB 61 PWLEKEIEATKKGKHPVIG 82  
XX  
XX RESULT 11  
XX ID AAM67139  
XX AC AAM67139;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27445.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX XX  
XX PM WO200157276-A2.  
XX PD 09-AUG-2001.  
XX XX

PF 30-JAN-2001; 2001WO-US00668.  
XX XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX DR WPI; 2001-488900/53.  
XX XX  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX PT analyzing gene expression in human bone marrow -  
XX XX  
XX PS Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC bone marrow. They can be used to measure gene expression in bone marrow  
XX CC samples, which may enable the improved diagnosis and treatment of cancers  
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX CC protein encoded by one of the probes of the invention.  
XX SO Sequence 82 AA:  
XX  
XX Query Match 14.1%; Score 438; DB 22; Length 82;  
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 280 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPVAMVWSQFVKYLIRPQ 339  
XX DB 1 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPVAMVWSQFVKYLIRPQ 60  
XX QY 340 PWLEKEIEATKKGKHPVIG 361  
XX DB 61 PWLEKEIEATKKGKHPVIG 82  
XX  
XX RESULT 12  
XX ID AAM15003  
XX AC AAM15003;  
XX DT 12-OCT-2001 (first entry)  
XX DE Peptide #1437 encoded by probe for measuring cervical gene expression.  
XX XX  
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;  
XX KW cervical cancer.  
XX OS Homo sapiens.  
XX XX  
XX PM WO200157278-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00670.  
XX XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI: 2001-488901/53.  
DR  
XX  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 27: SEQ ID No 19829; 487pp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
(SENPs: see A1110068-A128459). The present sequence is a peptide encoded  
by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
can be used to produce a single exon microarray, which can be used for  
measuring human gene expression in a sample derived from human cervical  
epithelial cells. By measuring gene expression, the probes are therefore  
useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 82 AA;  
Query Match 14.1%; Score 438; DB 22; Length 82;  
Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 280 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHVGDPVWWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHVGDPVWWSQFVKYLIRPQ 60  
QY 340 PWLEKEIEEATKKLGFKHPYIG 361  
DB 61 PWLEKEIEEATKKLGFKHPYIG 82  
DE Peptide #1477 encoded by probe for measuring placental gene expression.  
XX  
DE Probe: microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX Homo sapiens.  
XX  
XX W0200157272-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00663.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI: 2001-48897/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for  
PT

PT analyzing gene expression in human placenta -  
XX  
XX Claim 27: SEQ ID No 27709; 654pp; English.  
XX  
XX  
CC The present invention relates to single exon nucleic acid probes (SENPs:  
CC see A1131315-A157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 82 AA;  
Query Match 14.1%; Score 438; DB 22; Length 82;  
Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 280 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHVGDPVWWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHVGDPVWWSQFVKYLIRPQ 60  
QY 340 PWLEKEIEEATKKLGFKHPYIG 361  
DB 61 PWLEKEIEEATKKLGFKHPYIG 82  
DE Peptide #1412 encoded by probe for measuring breast gene expression.  
XX  
DE Probe: human; breast disease; breast cancer; development disorder;  
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
XX Homo sapiens.  
XX  
XX W0200157270-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 29-JAN-2001; 2001WO-US00661.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI: 2001-476286/51.  
DR  
XX Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -  
XX  
XX Claim 27: SEQ ID No 11470; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see A1100010-A110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic



CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 82 AA:

Query Match 14.1%; Score 438; DB 22; Length 82;  
Best Local Similarity 98.8%; Pred. No. 2.2e-34;  
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 280 EVKDKNVQVELPIVDSVHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 339  
Db 1 EVKDKNVQVELPIVDSVHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 60

Oy 340 PWLEKEIEATKKGFKHPVIG 361  
Db 61 PWLEKEIEATKKGFKHPVIG 82

RESULT 15

AAM64859 ID AAM64859 standard; Protein: 81 AA.

XX AC AAM64859;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36964.

XX KW Human: brain expressed exon; gene expression analysis; probe:  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI: 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX PT brains -

XX PS Example 4: SEQ ID NO: 36964; 650pp + Sequence Listing: English.

XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of  
XX CC the probes of the invention.

XX Sequence 81 AA:

Query Match 14.0%; Score 432; DB 22; Length 81;  
Best Local Similarity 98.8%; Pred. No. 8.3e-34;  
Matches 80; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 280 EVKDKNVQVELPIVDSVHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 339  
Db 1 EVKDKNVQVELPIVDSVHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 60

Oy 340 PWLEKEIEATKKGFKHPVI 360  
Db 61 PWLEKEIEATKKGFKHPVI 81

Search completed: October 31, 2002, 13:08:08  
Job time : 39 secs

**THIS PAGE BLANK (USPTO)**



181 AGGAGATGGCTGATCTCTCCGAATACGAGAGGCCCATTTGATGAGGGCCAGCTTCA 240  
181 AGGAGATGGCTGATCTCTCCGAATACGAGAGGCCCATTTGATGAGGGCCAGCTTCA 240  
241 GGAAGATTCGCTGCTTTAGAGAGCAATTTATGAGAGGCCCAAGACAGATTTGAAATTTAT 300  
241 GGAAGATTCGCTGCTTTAGAGAGCAATTTATGAGAGGCCCAAGACAGATTTGAAATTTAT 300  
301 AAGAAACAACTAAATGCTCCAGGAAAGATCATGAAATCTTAAGAGAGAGATTTGAA 360  
301 AAGAAACAACTAAATGCTCCAGGAAAGATCATGAAATCTTAAGAGAGAGATTTGAA 360  
361 AAGGAGCTTAAGAGCTTCGCTTTTCTCAAGTGTGAGTGAAGAAATTAAGAAATTTA 420  
361 AAGGAGCTTAAGAGCTTCGCTTTTCTCAAGTGTGAGTGAAGAAATTAAGAAATTTA 420  
421 GAAGAAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTTGGACATCATGAA 480  
421 GAAGAAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTTGGACATCATGAA 480  
481 AGGCTATATATGAGATCTATCTACTACCTCAAGCAAGATGGGCGAGGTGATGGCGT 540  
481 AGGCTATATATGAGATCTATCTACTACCTCAAGCAAGATGGGCGAGGTGATGGCGT 540  
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGAGAAATTAACATTTCTTCAGAAAT 600  
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGAGAAATTAACATTTCTTCAGAAAT 600  
601 CCCAAGGACTGCAGCAAGCCAAAGAGCTAGTGTATATCAACAAGAGCTGTGGCTAT 660  
601 CCCAAGGACTGCAGCAAGCCAAAGAGCTAGTGTATATCAACAAGAGCTGTGGCTAT 660  
661 GGCTGTACGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
661 GGCTGTACGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
721 CTGCGCTTGAATCTCACAATTTGGCGCTAGCTACTGAGGAGATGGAAGATGTTTGA 780  
721 CTGCGCTTGAATCTCACAATTTGGCGCTAGCTACTGAGGAGATGGAAGATGTTTGA 780  
781 CCGTAAAGTGAAGAGCTGCACAGACAGATCTGAGAGCTCCACTGAGACATTTGTCAGAGTGA 840  
781 CCGTAAAGTGAAGAGCTGCACAGACAGATCTGAGAGCTCCACTGAGACATTTGTCAGAGTGA 840  
841 GTAAAGGACAAATGTTGAGGTGAGTGGCTCCATTTAGACATGTTCCATCCGT 900  
841 GTAAAGGACAAATGTTGAGGTGAGTGGCTCCATTTAGACATGTTCCATCCGT 900  
901 CCTCATATTTACCCCTGCTGTCAGAGAGCTTGCAGATGACATTTGTCAGAGTCCAT 960  
901 CCTCATATTTACCCCTGCTGTCAGAGAGCTTGCAGATGACATTTGTCAGAGTCCAT 960  
901 CCTCATATTTACCCCTGCTGTCAGAGAGCTTGCAGATGACATTTGTCAGAGTCCAT 960  
901 CCTCATATTTACCCCTGCTGTCAGAGAGCTTGCAGATGACATTTGTCAGAGTCCAT 960  
961 GGTGATCTCTCAGTGTGAGTGTGATCCCACTTTGTCAAGTACTTGATCCGCCCAACCC 1020  
961 GGTGATCTCTCAGTGTGAGTGTGATCCCACTTTGTCAAGTACTTGATCCGCCCAACCC 1020  
1021 TGGCTGGAAGAGAAATGAGAAAGAGCCCAAGAGCTAGGCTTCAAAACATCCAGTTAT 1080  
1021 TGGCTGGAAGAGAAATGAGAAAGAGCCCAAGAGCTAGGCTTCAAAACATCCAGTTAT 1080  
1081 GAGATGCTATTTAGAGCAGACAGCAAGAGTGGAGCGAAGCAGCTTCCATCCATTTGAG 1140  
1081 GAGATGCTATTTAGAGCAGACAGCAAGAGTGGAGCGAAGCAGCTTCCATCCATTTGAG 1140  
1141 GAATPACAGGTGACGTTGAGAAAGACTTTTACGCTTCTGTCGACAGATGCAAGTGGAT 1200  
1141 GAATPACAGGTGACGTTGAGAAAGACTTTTACGCTTCTGTCGACAGATGCAAGTGGAT 1200  
1201 AAAAAAGGCTGATTTGGGACAGATGACCTGCTTGTAAAGAGGCAAAAAACAAG 1260  
1201 AAAAAAGGCTGATTTGGGACAGATGACCTGCTTGTAAAGAGGCAAAAAACAAG 1260

1261 TACCCAGTTATGAAATTTATTTAGTGAATTAACCTATCTCTTGTGACGTGACATCAATAT 1320  
1261 TACCCAGTTATGAAATTTATTTAGTGAATTAACCTATCTCTTGTGACGTGACATCAATAT 1320  
1321 CGATATACGAAATTTACCTTGGGGGTGATCCGATATTAACCTTCTCTCCAGGCA 1380  
1321 CGATATACGAAATTTACCTTGGGGGTGATCCGATATTAACCTTCTCTCCAGGCA 1380  
1381 GACTTCTAGTGTGATCTTTTTCATCTGAGAGTCTGTAAGTTCCTTATGAAATCATGCA 1440  
1381 GACTTCTAGTGTGATCTTTTTCATCTGAGAGTCTGTAAGTTCCTTATGAAATCATGCA 1440  
1441 GCGCTGATCTGATGCTGCTGCGAACTTCCTTTGATGACATCTATTTTGA 1500  
1441 GCGCTGATCTGATGCTGCTGCGAACTTCCTTTGATGACATCTATTTTGA 1500  
1501 GGGCCAAATGCCCAACCAATTTGCTATTCACCAACCTGGAATGAGAGAA 1560  
1501 GGGCCAAATGCCCAACCAATTTGCTATTCACCAACCTGGAATGAGAGAA 1560  
1561 ATCCCATGGAACCTGGAGATTTATTTGATGCTGGAATGACATGAGATGCTATCT 1620  
1561 ATCCCATGGAACCTGGAGATTTATTTGATGCTGGAATGACATGAGATGCTATCT 1620  
1621 AAGGTTTAAAGAAACTGGAAGAGCGGCTATATCTCTCTCAAAATTTGAGAG 1680  
1621 AAGGTTTAAAGAAACTGGAAGAGCGGCTATATCTCTCTCAAAATTTGAGAG 1680  
1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAATTA 1728  
1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAATTA 1728

## RESULT 2

US-09-839-136-9  
Sequence 9, Application US/09839136  
Patent No. US20020081694A1  
GENERAL INFORMATION:  
APPLICANT: Naoyuki TANIGUCHI et al.  
TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
FILE REFERENCE: 2356-7  
CURRENT APPLICATION NUMBER: US/09/839,136  
CURRENT FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 09/442,629  
PRIOR FILING DATE: 1999-11-18  
PRIOR APPLICATION NUMBER: 08/913,805  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: PCT/JP97/00171  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 2100  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (198)...(1925)  
US-09-839-136-9

Query Match 87.28; Score 1507.2; DB 10; Length 2100;  
Best local similarity 92.0%; Pred. No. 0;  
Matches 1590; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1 ATGCGGCATGAGACTGGTGTGGCGTTGGATTAATGCTCATTTCTTTGGCTGGGAGCC 60  
198 ATGCGGCATGAGACTGGTGTGGCGTTGGATTAATGCTCATTTCTTTGGCTGGGAGCC 257  
61 TTGCTATTTACATAGTGTGCTCACTTGTGAGAGATTAATGACACACTCTGACTTAGC 120  
258 TTGCTGTTTATATAGTGTGCTCACTTGTGAGAGATTAATGACACACTCTGACTTAGC 317  
121 CGAGACTGTCCAGATTTTGGCAAGAGCTGGAACGCTTTAAACACAAATGAAGACTTG 180

```
|||||
Db 318 CGAGACTGTCGAAGATTCGGCAAGCTTGAACGCTTAAACAGAGAAATGAAGACTTG 377
Oy 181 AGGAGAAATGGCTGATCTCTCCGAATACAGAAAGGCCCTTTATATAGGGCCAGCTTCA 240
Db 378 AGCGGAATGGCGCAATCTCTCCGATACCAAGAGCCCTATATGATGAGGGCCAGCTAT 437
Oy 241 GGAAGAGTTCGTCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTGAATTTAT 300
Db 438 GGAAGAGTACGCTTTTAGAGAGCAAGCTTTGTTAAGGCCAAGACAGATTGAATTTAC 437
Oy 301 AAGAAACAATAAAAAATGCTCCAGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 350
Db 498 AAGAAACAACCAAGAAATGGTCTGGGGAAGATCATGAATCTTGAAGAGAGATTGAA 557
Oy 361 AATGAGCTAAAGAGCTCTGCTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAAATTA 420
Db 558 AATGAGCTAAAGAGCTCTGCTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAAATTA 617
Oy 421 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTATCAGATTGGGACATCATGAA 480
Db 618 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTTATTTAGATTTAGACATCATGAA 677
Oy 481 AGGTCTATATGACGAGATCTATCTACTCTCAGTCAAAACAGATGGGAGAGTGTGCGT 540
Db 678 AGGTCTATATGACGAGATCTATCTACTCTCAGTCAAAACAGATGGGAGAGTGTGCGG 737
Oy 541 GAAAGAGGCCCAAGATCTGACAGAGCTGCTCAGCGGAGATTAACATATCTTCAGAA 600
Db 738 GAAAGAGGCCCAAGATCTGACAGAGCTGCTCAGCGGAGATTAACATATCTTCAGAA 797
Oy 601 CCCAAGAGCTGCGCAAGAGCAAGAGCTAGTGTGAATATCAAAAGGCTGTGCTAT 660
Db 798 CCCAAGAGCTGCGCAAGAGCAAGAGCTGTGTGAATATCAAAAGGCTGTGCTAT 857
Oy 661 GGCTGTACCTCCATCATCTAGTGTGTGTATGATTGCAATGAGCCAGCAGACA 720
Db 858 GGCTGTACCTCCATCATCTAGTGTGTGTATGATTGCAATGAGCCAGCAGACA 917
Oy 721 CTCGCTTGAATCTCAAAATGGCCCTACGCTACCTGCGGAGATGGAACTGTGTTAGA 780
Db 918 CTCATCTTGAATCTCAAAATGGCCCTACGCTACCTGCGGAGATGGAACTGTGTTAGA 977
Oy 781 CCTGTAGTGAAGCTGACAGACAGATCTGAGCTCCACAGATGGAGTGTGAGGAA 840
Db 978 CCTGTAGTGAAGCTGACAGACAGATCTGAGCTCCACAGATGGAGTGTGAGGAA 1037
Oy 841 GTAAAGACAAAAATGTTGAGTGTGAGCTCCCATTTGTAGACAGTGTCTCCCTG 900
Db 1038 GTAAAGACAAAAATGTTGAGTGTGAGCTCCCATTTGTAGACAGTGTCTCCCTG 1097
Oy 901 CCTCCATATTTTACCTTGGCTGTCCAGAAAGACCTTGCAGATGACCTTGTAGATGCAT 960
Db 1098 CCTCCATATTTTACCTTGGCTGTGACAAAGACCTTGCAGATGACCTTGTAGATGCAT 1157
Oy 961 GGTGATCTGCAAGTGTGAGTGTGATCCAGTTGTCAAGTATGTTGCCCAACAACC 1020
Db 1158 GGTGATCTGCAAGTGTGAGTGTGATCCAGTTGTCAAGTATGTTGCCCAACAACC 1217
Oy 1021 TGGCTGAAAGAGAAATAGAAAGAGCCACCAAGAACTAGCTTCAAAACATTCAGTTAT 1080
Db 1218 TGGCTGAAAGAGAAATAGAAAGAGCCACCAAGAACTAGCTTCAAAACATTCAGTTAT 1277
Oy 1081 GGAAGTCAATGTTAGACGACAGACAAAGTGGAGGAGGAGCAGCTTCCATCCATTTAG 1140
Db 1278 GGAAGTCAATGTTAGACGACAGACAAAGTGGAGGAGGAGCAGCTTCCATCCATTTAG 1337
Oy 1141 GAATACAGGTGACGTGAGAGAACTTACCTTCTGCTGAGAAATGAAGAGTGGAT 1200
Db 1338 GAATACAGGTGACGTGAGAGAACTTACCTTCTGCTGAGAAATGAAGAGTGGAT 1397
Oy 1201 AAAAAAGGTGTATTTGGCCACAGATGACCTTGTGTTAAAGAGGCAAAAGCAAG 1260
|||||
```

---

```
Db 1398 AAAAAAGACTGTATTTGGCCACAGATGACCTTCTTATTTAAAGAGGCAAAACAAAG 1457
Oy 1261 TACCCAGTTATTAATTTATAGTGAATACCTATCTTGTGACAGCTGACATCATAT 1320
Db 1458 TACCCAGTTATTAATTTATAGTGAATACCTATCTTGTGACAGCTGACATCATAT 1517
Oy 1321 CGATATACAGAAATTTACCTTGGGAGTGTGATCTGATATACACTTTCTCTCCAGCA 1380
Db 1518 CGATATACAGAAATTTACCTTGGGAGTGTGATCTGATATACACTTTCTCTCCAGCA 1577
Oy 1381 GACTTCCTAGTGTATCTTTTATGATGAGTGTGATGATGATGATGATGATGATGATGAT 1440
Db 1578 GACTTCCTAGTGTATCTTTTATGATGAGTGTGATGATGATGATGATGATGATGATGAT 1637
Oy 1441 GCGCTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1638 ACATCTACATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697
Oy 1501 GGCCCAATGCGCCACAAACCAATTTGCAATTTATCTTATCTTATCTTATCTTATCTTAT 1560
Db 1698 GGCCCAATGCGCCACAAACCAATTTGCAATTTATCTTATCTTATCTTATCTTATCTTAT 1757
Oy 1561 ATCCCATGGAACCTGAGATATTTATTTGATGATGATGATGATGATGATGATGATGAT 1620
Db 1758 ATCCCATGGAACCTGAGATATTTATTTGATGATGATGATGATGATGATGATGATGAT 1817
Oy 1621 AAGGTGTTTACGAAATCTGGAGAGAGCGGCTATATCTCTCTTCAAAAGTTCCAGAG 1680
Db 1818 AAGGTGTTTACGAAATCTGGAGAGAGCGGCTATATCTCTCTTCAAAAGTTCCAGAG 1877
Oy 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAGTAA 1728
Db 1878 AAGATGAAACAGTCAAGTACCCACATATCTGAGGCTGAGAAATTA 1925
```

---

```
RESULT 3
US-09-864-761-13359
; Sequence 13359, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13359
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
US-09-864-761-13359
```

```
Query Match          15.9%; Score 275.2; DB 10; Length 503;
Best Local Similarity 91.2%; Pred. No. 5.5e-71;
Matches 292; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
```

```
QY 1409 AGGTCTGTAGAGTTGCTTATGAAATCATGCAACCGCTGCTGATGCTTGGCAACT 1468
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 9 AGGTCTGTAGAGTTGCTTATGAAATCATGCAACCGCTGCTGATGCTTGGCAACT 68
QY 1469 TCCGTTCTTGGATGACATCTACTATTTTGGAGGCCCAATGCCACCAACCAATTGCCA 1528
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 69 TCCATCTCTTACATGACATCTACTATTTTGGGGCCGAGATGCCCAATCAATTCGCCA 128
QY 1529 TTTATCTCTACCAACCTGACAGTGAAGAGAAATCCCATGGAACCTGGAGATATTATTG 1588
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 129 TTTATCTCTACCAACCGGAGATGAGATGAAATTCCTGGAACCTGGAGATATTATTG 188
QY 1589 GTTGTGCTGAATTCACATGGGATGGCTATCTTAAAGTGTTAACAGAAACTGGGAAGCA 1648
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 189 GTTGTGCTGAATTCACATGGGATGGCTATCTTAAAGTGTCAACAGAAATTTGGGAAGCA 248
QY 1649 CGGGCCATATCTCCTCTACAAAGTTGAGAGAAATAGAAAGTGAAGTACCCCAT 1708
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 249 CGGGCCATATCTCCTCTACAAAGTTGAGAGAAATAGAAAGTGAAGTACCCCAT 308
QY 1709 ATCCCGAGCTGACAACTAA 1728
      ||||| ||||||| |||||
DB 309 ATCTGAGGCTGAGAAATAA 328
```

```
RESULT 4
US-09-864-761-13292
; Sequence 13292, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
```

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13292
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-13292
```

```
Query Match          12.7%; Score 218.8; DB 10; Length 551;
Best Local Similarity 91.3%; Pred. No. 2.1e-54;
Matches 232; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 829 TGTGACAGTGAAGTAAGCAAAATGTTACAGTGTGTTGACCTCCCATTTGACAGT 888
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 172 TGTGACAGTGAAGTAAGCAAAATGTTACAGTGTGTTGACAGTGTGTTGACAGT 231
QY 889 GTTCATCTCTGCTCCATATTATACCCCTGGCTGTCCAGAAAGCTTGCAGATCGACTT 948
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 232 GTTCATCTCTGCTCCATATTATACCCCTGGCTGTCCAGAAAGCTTGCAGATCGACTT 291
QY 949 GTACAGTCCATGTGTATCTGTCAGTGTGGTATCCAGTTGTCAAGTACTTGATT 1008
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 292 GTACAGTCCATGTGTATCTGTCAGTGTGGTATCCAGTTGTCAAGTACTTGATT 351
QY 1009 CGCCCAACACCTTGGCTGGGAAAGAAATAGAAAGAGCCCAAGAGCTTCAAA 1068
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 352 CGCCCAACACCTTGGCTGGGAAAGAAATAGAAAGAGCCCAAGAGCTTCAAA 411
QY 1069 CATCCAGTTATTGG 1082
      ||||||| |||||||
DB 412 CATCCAGTTATTGG 425
```

```
RESULT 5
US-09-864-761-18181
; Sequence 18181, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
```

```

APPLICANT: Rank, David R.
APPLICANT: Hanzel, Jensheng
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18181
LENGTH: 248
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL109847.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
OTHER INFORMATION: SWISSPROT HIT: C9YL7, EVALU0 6.50e+00
OTHER INFORMATION: NT HIT: AF036280.1, EVALU0 0.00e+00
OTHER INFORMATION: EST-HUMAN HIT: AM387766.1, EVALU0 0.00e+00
US-09-864-761-18181

```

	Query Match	12.3%	Score 213.4	DB 10	Length 248
	Best Local Similarity	91.5%	Pred. No. 5e-53		
	Matches 226	Conservative 0	Mismatches 21	Indels 0	Gaps 0
Oy	836 GTGAGCTAAGACAAATAATGCTTACAGTCGCTTGACATCCCAATTGACAGACGTGTCATCC				895

Db	1	GTGAGTGTAAGSACAAAATGTTCAGATGCTGCAGCTTCCATTTGTAGACAGTCTTATC	60
QY	896	CTGCTCTCCATATTTTACCCCTGGCTGTCAGAAAGACCTTGCAGATGACTTGTACGAG	955
Db	61	CCCGCTCTCATATTTTACCTTGGCTGTACCAAGAACCTTGCAGATGACTTGTACGAG	120
QY	956	TCCATGCTGATCTGCACTGTGTGGTATCCCACTTTGTCAACTACTTGAATTTGCCAC	1015
Db	121	TGCATGTGAGCCCTGCATGTGTGGTGTCTCAGTTTGTCAAACTACTTGAATTCGCCAC	180
QY	1016	AACCTTGGCTGGAAGAAATAGAGAGGACACCAAGAACCTTGGCTTCAACATCTCAG	1075
Db	181	AGCCTTGGCTGGAAGAAATAGAGAGGACACCAAGAACCTTGGCTTCAACATCTCAG	240
QY	1076	TTATTTGG 1082	
Db	241	TTATTTGG 247	
RESULT 6			
US-09-864-761-30284			
: Sequence 30284, Application us/09864761			
: Patent No. US20020048763A1			
GENERAL INFORMATION:			
APPLICANT: Penn, Sharron G.			
APPLICANT: Rank, David R.			
APPLICANT: Hanzel, David K.			
APPLICANT: Chen, Wenshang			
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR			
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY			
FILE REFERENCE: Aeomica-X-1			
CURRENT APPLICATION NUMBER: US/09/864, 761			
CURRENT FILING DATE: 2001-05-23			
PRIOR APPLICATION NUMBER: US 60/180,312			
PRIOR FILING DATE: 2000-02-04			
PRIOR APPLICATION NUMBER: US 60/207,456			
PRIOR FILING DATE: 2000-05-26			
PRIOR APPLICATION NUMBER: US 09/632,366			
PRIOR FILING DATE: 2000-08-03			
PRIOR APPLICATION NUMBER: GB 24263. 6			
PRIOR FILING DATE: 2000-10-04			
PRIOR APPLICATION NUMBER: US 60/236,359			
PRIOR FILING DATE: 2000-09-27			
PRIOR APPLICATION NUMBER: PCT/US01/00666			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00667			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00664			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00669			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00665			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00668			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00663			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00662			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00661			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: PCT/US01/00670			
PRIOR FILING DATE: 2001-01-30			
PRIOR APPLICATION NUMBER: US 60/234,687			
PRIOR FILING DATE: 2000-09-21			
PRIOR APPLICATION NUMBER: US 09/608,408			
PRIOR FILING DATE: 2000-06-30			
PRIOR APPLICATION NUMBER: US 09/774,203			
PRIOR FILING DATE: 2001-01-29			
NUMBER OF SEQ ID NOS: 49117			
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1			
SEQ ID NO 30284			
LENGTH: 248			

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q9VL77, EVALU6 6.50e+00
; OTHER INFORMATION: EST_HUMAN HIT: AN387766.1, EVALU6 0.00e+00
; OTHER INFORMATION: NT HIT: AF038280.1, EVALU6 0.00e+00
US-09-864-761-30284
```

```

Query Match      12.3%: Score 213.4; DB 10; Length 248;
Best Local Similarity 91.5%: Pred. No. 5e-53;
Matches 226; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```

QY 836 GTGAGTAAAGACAAATATGTCAGGTGGTGTGCTCCCATTTGATGATGTTATC 895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 GTGAGTAAAGACAAATATGTCAGGTGGTGTGCTCCCATTTGATGATGTTATC 60
QY 896 CTGCTCCCATATTTTACCCCTGGCTGCCAGAGACTTGCAGATCGACTGTACGAG 955
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 CCGGTCCCATATTTTACCCCTGGCTGCCAGAGACTTGCAGATCGACTGTACGAG 120
QY 956 TCATGTGTCTCTGCGAGTGTGGTATCCAGTTTGTCAAGTACTTATTCGCCAC 1015
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 TCGATGTGACCTCGAGTGTGGTGTCTCAGTTTGTCAATACTTGTATCCGCCAC 180
QY 1016 ACCCTGTGAAAGAAATAGAGAGGCGACCAAGAGCTAGGCTCAACATCCAG 1075
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 AGCTTGTGCTAGAAAGAAATAGAGAGGCGACCAAGAGCTTGGCTCAACATCCAG 240
QY 1076 TTATTTG 1082
    |||||
DB 241 TTATTTG 247
```

## RESULT 7

```

; Sequence 1421, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1421
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
US-09-864-761-1421
```

```

Query Match      8.1%: Score 139.6; DB 10; Length 384;
Best Local Similarity 91.4%: Pred. No. 3.2e-31;
Matches 148; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```

QY 829 TGTGAGTGAAGTAAGACAAATATGTCAGGTGGTGTGCTCCCATTTGACAGT 888
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 223 TTGTGAGTGAAGTGAAGACAAATATGTCAGGTGGTGTGCTCCCATTTGACAGT 282
QY 889 GTTCATCTCTGCTCCATATTTTACCCCTGGCTGCCAGAGACTTGCAGATCGACTT 948
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 283 CTTCATCTCTGCTCCATATTTTACCCCTGGCTGCCAGAGACTTGCAGATCGACTT 342
QY 949 GTGAGTGCATGTGATCTCGAGTGTGGTGTGCTCCCATTTGACAGT 990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 343 GTGAGTGCATGTGATCTCGAGTGTGGTGTGCTCCCATTTGACAGT 384
```

## RESULT 8

```

; Sequence 29920, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
```



```

: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 417
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
US-09-351-794A-3

Query Match          2.2%: Score 38.2; DB 10; Length 417;
Best Local Similarity 47.7%: Pred. No. 0.2;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

OY 258 AGAAGAGCAATTTATGAAGGCCAAGAAGACAGATTCGAATTAATATAGAACCAACTAATAAA 317
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 AGAAGAGTTAAATAGGATTTGAACAAGAAAAATTTGAATATGTCAATTAAGAAAGAAATAAA 242

OY 318 TGGTTCAGAGGAAGCATCTGAATTCCTAAGGAGAGAGATTGAAATGCGAGCTTAAGAGCT 377
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 TAAAMCTTACAAAGACGCAATTAATTAATATGTGATGTGATCAAGTTGATTAATATCATAG 302

OY 378 CTGCTTTTTTCTACAAAGCTGAGTTGAAGAATTTAAAGAAATTTAGAAGGAATGAACCTCA 437
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 303 AGAACAAATTAGAAAAATTCGAAAAAGAAAAAATTAACAAATGTGATTAAGAAATGAAATAGA 362

OY 438 AAGACATTCACATGATTTCTATCGATTTGGGACATCATGAAGCTTATTAATG 492
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 TAAATATATATAGAGAAAGAAATTAGACAAATGGATCGCATGCAATTTATAGTATG 417

RESULT 10
US-09-351-794A-1
: Sequence 1, Application US/09351794A
: Patent No. US20020042382A1
: GENERAL INFORMATION:
: APPLICANT: DUFFY, PATRICK E.
: APPLICANT: OCKENHOUSE, CHRISTIAN F.
: TITLE OF INVENTION: SEQUESTIN
: FILE REFERENCE: 38644-1/7519
: CURRENT APPLICATION NUMBER: US/09/351,794A
: PRIOR FILING DATE: 1999-07-13
: PRIOR APPLICATION NUMBER: 08/559,896
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 1956
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
US-09-351-794A-1

Query Match          2.2%: Score 38.2; DB 10; Length 1956;
Best Local Similarity 47.7%: Pred. No. 0.54;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

OY 258 AGAAGAGCAATTTATGAAGGCCAAGAAGACAGATTCGAATTAATATAGAACCAACTAATAAA 317
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 306 AGAAGAGTTAAATAGGATTTGAACAAGAAAAATTTGAATATGTCAATTAAGAAAGAAATAAA 365

OY 318 TGGTTCAGAGGAAGCATCTGAATTCCTAAGGAGAGGATTTGAAATGAGAGCTTAAGAGCT 377
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 TAAACCTTACAAGGCGCAATTAATTAATATGATGATCAAGTTGATTAATAATTGATAG 425

OY 378 CTGCTTTTTTCTACAAAGCTGAGTTGAAGAATTTAAGAAATTTAGAACGAATGAACCTCA 437
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 AGAAGCAATTTAGAAAAATTCGAAAAAGAAAAAATTAACAAATGTGATTAAGATGAATAGA 485

OY 438 AAGACATTCAGATGATTTCTATCGATTTGGGACATCATGAAGGCTTATTAATG 492
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 486 TAAATATATATAGAGAAAGAAATTAGACAAATGGATCGCATGCAATTTATAGTATG 540

```

```
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalaigan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3543
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 38-LIB3058-050-Q1-K1-B10
US-09-983-965-3543
```

```
Query Match          2.1%; Score 35.8; DB 10; Length 442;
Best Local Similarity 49.2%; Pred. No. 1.1;
Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
```

```
QY 324 AGGAGAGATCATGAATCTCTAAGAGAGAGATTTGAAATGACCTAAGACCTGCTT 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 AGGCGACATTTAGAGTAGTATGAGAACTGAGAGTATGATTCAGAGAAATTAAGATGGC 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 384 TTTTCTACAAAGTGTGAAGAAATTAAGAAATTTAGAAGAAATGCACTCCAAAGACA 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 AATTAATACAAAGAGTAAGAAATTTCTAATATTTCTAAGAGAAATATTTCTAAGAAA 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 444 TGCAGATGAATTTCTATCAGATTTGGACATCATGAAGGCTCTATATGACGATCTATA 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 TGAATAATTTAATGCAATTAAGAGGTGATTTCTATCTAATCTATCAATGAGGATTTTTC 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 504 CTACCTCAGTC 514
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 AAGCCTCAATC 112
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 12

```
; Sequence 3291, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3291
; LENGTH: 32189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3291
```

```
Query Match          2.1%; Score 35.6; DB 10; Length 32189;
Best Local Similarity 51.2%; Pred. No. 18;
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
```

```
QY 280 AAGACACAGATTTTAAAGAAACAATAAAATGTCGAGGAGATCATGAA 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23875 ATGAGATGCATTTGAGGTTCTAATAATAAAGATTAATTAACGTAAGGACTAA 23934
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 340 ATCCTAAGAGAGATTTGAAATGAGCTAAGAGCTGTGTTTCTTCAAAAGTGAG 399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 23935 ATCGGGGAGAGAAAGATTGCAATTAATACATGGCCCAAGATGTTGGTCCCATGG 23994
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 TTGAAGAAATTAAGATTTAGAAGAAATGAACCTCCAAAGA 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23995 AGTTTAAATTAATAAATAATTAAGAAACAAATACCCCAA 24036
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 13

```
US-09-960-352-12911/C
; Sequence 12911, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalaigan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12911
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 55-LIB3058-057-Q1-K1-F12
US-09-960-352-12911
```

```
Query Match          2.0%; Score 35.2; DB 10; Length 442;
Best Local Similarity 49.5%; Pred. No. 1.6;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

```
QY 260 AAGACATTTATGAGAGCCCAAGACAGATTGAAAATTTAAGAAACAACATAAATG 319
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 370 AAAAAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 311
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 320 GTCCAGGAGAGATTCATGAATCTTAAGAGAGATTTGAATTTGAGTAAGAGCTCT 379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 310 AAAAAAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 251
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 GGTTTTCTACAAAGTGTGAGCAATTTAAGAAATTTAAGAGAAATGAACTCCAA 439
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 250 CAATTAATAATATATTAATAAATAAACAATAAATAAATAAATAAATAAATAAATAA 191
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 440 GACA 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 190 AAAA 187
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 14

```
US-09-960-352-5087/C
; Sequence 5087, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalaigan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5087
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 22-LIB3058-032-Q1-K1-F9
US-09-960-352-5087
```

```
Query Match          2.0%; Score 34.8; DB 10; Length 376;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 260 AAGAGCAATTTATGAGGCCAAGACAGATTGAAAATTATAGAAACAACCTAAAAATG 319
DB 374 AAAAAATATAAAAAAAGAAAAATATAAAAAATATAACAAAATTAACAATAATATA 315
QY 320 GTCCAGGAGGATCATGAATCTCTAAGAGAGCATTTGAAATGAGCTAAAGACCTCT 379
DB 314 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 255
QY 380 GGTCTTTCTACAAAGTGAAGATTGAAGAAATTAGAAGAAATGAATCCCAA 439
DB 254 AATTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 195
QY 440 GA 441
DB 194 AA 193

RESULT 15
US-09-759-841-3
; Sequence 3, Application US/09759841
; Patent No. US20010039026A1
; GENERAL INFORMATION:
; APPLICANT: Rickett, Graham A
; APPLICANT: Dobbs, Susan
; APPLICANT: Perros, Manousos
; TITLE OF INVENTION: Assay Method
; FILE REFERENCE: PC10348APME
; CURRENT APPLICATION NUMBER: US/09/759,841
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: GB 0000661.9
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: GB 0000663.5
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: GB 0000659.3
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Human Immunodeficiency virus type 1
US-09-759-841-3

Query Match          2.0%; Score 34.8; DB 10; Length 1512;
Best Local Similarity 50.6%; Pred. No. 4.6;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 254 CTTTAGAGCAATTTATGAGGCCAAGACAGATTGAAAATTATAGAAACAACCTA 313
DB 1013 CTTTAATATAGATAGTATATAAATTAGAGACAATTTGGAAATAAACAATAGTCTTA 1072
QY 314 AAAAAATGTCAGGAGGAGATCATGAATCTCTAAGAGAGAGATTTGAAAATGAGCTAAG 373
DB 1073 AGCATTCTCAGGAGGAGGAGCCAGAAATTTGTGACGCACAGTTTAAATTTGAGAGGGGAAT 1132
QY 374 AGCTCTGTTTCTTCTACAAAGTGAAGTTGAAGAAATTAAGAATTT 419
DB 1133 TTTTCTACTTAATTCACACAACACTGTTTAATAGTACTTGAAATGT 1178
```

Search completed: November 2, 2002, 00:06:17  
Job time : 118 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 21:34:37 : Search time 60 seconds  
(without alignments)  
8832.291 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728  
Sequence: 1 atgcggcatgactgctgtc.....atcccgagctgacgaatga 1728

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptcdat1/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptcdat1/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptcdat1/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptcdat1/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptcdat1/1/lna/6CTUS.COMB.seq:\*  
6: /cgn2\_6/ptcdat1/1/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1726.4	99.9	1728	3	US-08-913-805A-1	Sequence 1, Appl1
2	1726.4	99.9	1728	4	US-09-442-629-1	Sequence 1, Appl1
3	1505.6	87.1	2100	3	US-08-913-805A-9	Sequence 9, Appl1
4	1505.6	87.1	2100	4	US-09-442-629-9	Sequence 9, Appl1
5	75.6	4.4	7218	1	US-08-232-463-14	Sequence 14, Appl1
6	38.2	2.2	417	4	US-08-559-896B-3	Sequence 3, Appl1
7	38.2	2.2	1956	4	US-08-559-896B-1	Sequence 1, Appl1
8	37.8	2.2	1448	2	US-08-037-816A-19	Sequence 19, Appl1
9	37.8	2.2	1448	2	US-08-037-816A-23	Sequence 23, Appl1
10	37.8	2.2	1448	2	US-08-530-146-19	Sequence 19, Appl1
11	37.8	2.2	1448	2	US-08-530-146-23	Sequence 23, Appl1
12	37.8	2.2	1532	2	US-08-037-816A-15	Sequence 15, Appl1
13	37.8	2.2	1532	2	US-08-037-816A-27	Sequence 27, Appl1
14	37.8	2.2	1532	2	US-08-530-146-15	Sequence 15, Appl1
15	37.8	2.2	1532	2	US-08-530-146-27	Sequence 27, Appl1
16	36.8	2.1	4527	2	US-08-944-449-8	Sequence 8, Appl1
17	36.8	2.1	4527	4	US-09-353-362-8	Sequence 8, Appl1
18	36.4	2.1	2960	3	US-08-913-842-3	Sequence 3, Appl1
19	36	2.1	1525	1	US-08-609-572-1	Sequence 1, Appl1
20	36	2.1	1525	4	US-08-841-751-1	Sequence 1, Appl1
21	36	2.1	1525	4	US-08-846-340-1	Sequence 1, Appl1
22	36	2.1	1525	4	US-08-846-344-1	Sequence 1, Appl1
23	35.8	2.1	1461	3	US-08-889-841B-7	Sequence 7, Appl1
24	35.2	2.0	162450	4	US-09-345-882-1	Sequence 1, Appl1
25	35	2.0	14176	1	US-08-307-499-1	Sequence 1, Appl1
26	35	2.0	14176	1	US-08-307-499-14	Sequence 14, Appl1
27	35	2.0	14176	4	US-09-299-268-1	Sequence 1, Appl1

28	35	2.0	14176	4	US-09-299-268-14	Sequence 14, Appl1
29	34.6	2.0	675	1	US-08-307-499-55	Sequence 55, Appl1
30	34.6	2.0	675	4	US-09-299-268-55	Sequence 55, Appl1
31	34.6	2.0	1503	3	US-08-889-841B-1	Sequence 1, Appl1
32	34.6	2.0	1503	3	US-08-889-841B-4	Sequence 4, Appl1
33	34.6	2.0	1539	4	US-07-956-483-21	Sequence 21, Appl1
34	34.6	2.0	1539	4	US-08-472-240A-13	Sequence 13, Appl1
35	34.6	2.0	1759	1	US-08-105-483-279	Sequence 279, App
36	34.6	2.0	1759	1	US-08-709-209-279	Sequence 279, App
37	34.6	2.0	1759	1	US-08-458-101-279	Sequence 279, App
38	34.6	2.0	2020	2	US-08-417-210A-135	Sequence 135, App
39	34.6	2.0	2028	2	US-08-417-210A-138	Sequence 138, App
40	34.6	2.0	2060	2	US-08-417-210A-141	Sequence 141, App
41	34.6	2.0	2571	1	US-08-254-358-3	Sequence 3, Appl1
42	34.6	2.0	2571	1	US-08-475-391-3	Sequence 3, Appl1
43	34.6	2.0	2571	2	US-08-709-609-3	Sequence 3, Appl1
44	34.6	2.0	2571	5	PCT-US95-07178-3	Sequence 3, Appl1
45	34.6	2.0	3807	2	US-08-417-210A-78	Sequence 78, Appl1

## ALIGNMENTS

RESULT 1  
US-08-913-805A-1  
Sequence 1, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naofumi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YAMAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
City: Washington  
State: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetel, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-913-805A-1

Query Match 99.98; Score 1726.4; DB 3; Length 1728;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ATCGGCGCATGAGCTGGTCTGGCGCTTGATATGCTCATTTCTTTTGGCTGGGGACC 60
    |||||||
Db 1 ATCGGCGCATGAGCTGGTCTGGCGCTTGATATGCTCATTTCTTTTGGCTGGGGACC 60

QY 61 TTGCTATTTTACATAGGTGTCTCCTGGTACGAGATATGACCACTCTGATCCTAGC 120
    |||||||
Db 61 TTGCTATTTTACATAGGTGTCTCCTGGTACGAGATATGACCACTCTGATCCTAGC 120

QY 121 CGAAGACTGTCAGATTTTGGCAAGCTGGAACGCTTAAACACAAATGGAAGACTTG 180
    |||||||
Db 121 CGAAGACTGTCAGATTTTGGCAAGCTGGAACGCTTAAACACAAATGGAAGACTTG 180

QY 181 AGGAGAAATGCTGATCTCTCCGAATACGAAAGGCCCATTTGTCAGGGGCCAGCTTCA 240
    |||||||
Db 181 AGGAGAAATGCTGATCTCTCCGAATACGAAAGGCCCATTTGTCAGGGGCCAGCTTCA 240

QY 241 GGAAGAGTTCTGCTTTAGAAAGCAATTTATGAAGGCCAAAGACAGATTGAAATAT 300
    |||||||
Db 241 GGAAGAGTTCTGCTTTAGAAAGCAATTTATGAAGGCCAAAGACAGATTGAAATAT 300

QY 301 AAGAAACAACTAAAAATGCTCCAGGAGAGATCATGAATTCCTAAGGAGAGATTGAA 360
    |||||||
Db 301 AAGAAACAACTAAAAATGCTCCAGGAGAGATCATGAATTCCTAAGGAGAGATTGAA 360

QY 361 AATGAGACTAAGAGCTCTGTTTCTCAAAAGTGTGAAGAAATTAAGAAATTTA 420
    |||||||
Db 361 AATGAGACTAAGAGCTCTGTTTCTCAAAAGTGTGAAGAAATTAAGAAATTTA 420

QY 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCAT 480
    |||||||
Db 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCAT 480

QY 481 AGGTCTATATGAGATCTATCTACTACCTGATCAACCAATGAGGAGGTGATGGCGT 540
    |||||||
Db 481 AGGTCTATATGAGATCTATCTACTACCTGATCAACCAATGAGGAGGTGATGGCGT 540

QY 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGGAGAAATATCATTTTCAAGAT 600
    |||||||
Db 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGGAGAAATATCATTTTCAAGAT 600

QY 601 CCCAAGACTGACGACAAAGCCAAAGACTAGTGTATATCAACAAAGGCTGTGCTAT 660
    |||||||
Db 601 CCCAAGACTGACGACAAAGCCAAAGACTAGTGTATATCAACAAAGGCTGTGCTAT 660

QY 661 GGTGTACGTCTCATCATATGATGTAAGTACTGCTTATGATTTGATTCACATTCGCCACCGACGA 720
    |||||||
Db 661 GGTGTACGTCTCATCATATGATGTAAGTACTGCTTATGATTTGATTCACATTCGCCACCGACGA 720

QY 721 CTGCGCTTGAATCTCATATTTGCGCTACGCTACTGAGGAGATGGAATGATTTTGA 780
    |||||||
Db 721 CTGCGCTTGAATCTCATATTTGCGCTACGCTACTGAGGAGATGGAATGATTTTGA 780

QY 781 CCTGTAAAGTGAAGCTGACACAGACAGATCTGAGAGCTCCACTGAGACATTTGGTCAAGGTGAA 840
    |||||||
Db 781 CCTGTAAAGTGAAGCTGACACAGACAGATCTGAGAGCTCCACTGAGACATTTGGTCAAGGTGAA 840

QY 841 GTTAAAGGACAAAATGTTCAAGGTGTTGAGCTCCCATTTTACAGAGTTCATCTCGT 900
    |||||||
Db 841 GTTAAAGGACAAAATGTTCAAGGTGTTGAGCTCCCATTTTACAGAGTTCATCTCGT 900

QY 901 CCTTCATATTTACCCCGCTGTCAGGAGACCTTGACATGACTTGTACGAGTTCAT 960
    |||||||
Db 901 CCTTCATATTTACCCCGCTGTCAGGAGACCTTGACATGACTTGTACGAGTTCAT 960

QY 961 GGTGATCTGTCAGTGTGTGGGTATCCCAAGTTGTCAAGTACTTGATTCGCCACCAACC 1020
    |||||||

```

```

Db 961 GGTGATCTGTCAGTGTGTGGGTATCCCAAGTTGTCAAGTACTTGATTCGCCACCAACC 1020
    |||||||
QY 1021 TTGGCTGAAAAGGAAATAGAAAGGCCACCAAGAGCTTAGGCTTCAACATTCAGTTATT 1080
    |||||||
Db 1021 TTGGCTGAAAAGGAAATAGAAAGGCCACCAAGAGCTTAGGCTTCAACATTCAGTTATT 1080

QY 1081 GGAATCCATGTTAGACGACACAGACAAAGTGGAGCGGAGACAGCTTCCATCCATTGAG 1140
    |||||||
Db 1081 GGAATCCATGTTAGACGACACAGACAAAGTGGAGCGGAGACAGCTTCCATCCATTGAG 1140

QY 1141 GAATACAGCTGACGCTTGAAGAGACTTTGAGTTCCTGCGACATATGACATGAGAT 1200
    |||||||
Db 1141 GAATACAGCTGACGCTTGAAGAGACTTTGAGTTCCTGCGACATATGACATGAGAT 1200

QY 1201 AAAAAAAGGTGTATTGGCCACAGATGACCTGCTTTGTTAAAGGGCAAAACAAAG 1260
    |||||||
Db 1201 AAAAAAAGGTGTATTGGCCACAGATGACCTGCTTTGTTAAAGGGCAAAACAAAG 1260

QY 1261 TACCCAGTTATGATTTATAGTATGATACCTATCTCTTGGTACGCTGACTACATAT 1320
    |||||||
Db 1261 TACCCAGTTATGATTTATAGTATGATACCTATCTCTTGGTACGCTGACTACATAT 1320

QY 1321 CGATATACAGAAATTCATCTTGGGGTGTGATCCTGATATACACTTTCCTCCAGCA 1380
    |||||||
Db 1321 CGATATACAGAAATTCATCTTGGGGTGTGATCCTGATATACACTTTCCTCCAGCA 1380

QY 1381 GACTTCTAGTGTACTTTTTCATGCGAGCTGTGATGATGCTTGTAAATCATGCAA 1440
    |||||||
Db 1381 GACTTCTAGTGTACTTTTTCATGCGAGCTGTGATGATGCTTGTAAATCATGCAA 1440

QY 1441 GCGCTGATCTGATGCTGCTGCGAATCTCGTTCTTTGGATGACATCTATTTTGA 1500
    |||||||
Db 1441 GCGCTGATCTGATGCTGCTGCGAATCTCGTTCTTTGGATGACATCTATTTTGA 1500

QY 1501 GCGCCAAATGCCACAAACAAATTCCTATTCCTCACCACCTGAGACTGAAGGAA 1560
    |||||||
Db 1501 GCGCCAAATGCCACAAACAAATTCCTATTCCTCACCACCTGAGACTGAAGGAA 1560

QY 1561 ATCCCATGAGACCTGAGATATTATTTGTTGTTGCTGGAATCAGTGGATGGCTATCCT 1620
    |||||||
Db 1561 ATCCCATGAGACCTGAGATATTATTTGTTGTTGCTGGAATCAGTGGATGGCTATCCT 1620

QY 1621 AAGGTGTTAAAGAAACCTGGGAGAGCGGCTATATTCCTTCAAAAGTTGAGAG 1680
    |||||||
Db 1621 AAGGTGTTAAAGAAACCTGGGAGAGCGGCTATATTCCTTCAAAAGTTGAGAG 1680

QY 1681 AAGTGAACACAGTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728
    |||||||
Db 1681 AAGTGAACACAGTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728

RESULT 2
US-09-442-629-1
; Sequence 1, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; UOZUMI, Naofumi
; SHIBA, Tetsuo
; YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-No. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetel, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-442-629-1

Query Match 99.9%; Score 1726.4; DB 4; Length 1728;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGCATGAGCTGTTGCGGCTGGATTATGCTATCTTTTTCCTGGGGACC 60  
DB 1 ATGCGGCATGAGCTGTTGCGGCTGGATTATGCTATCTTTTTCCTGGGGACC 60  
QY 61 TTGCTATTACATAGAGTGTGCTGCTGATGAGATATGACCACTGTGACTGTAGC 120  
DB 61 TTGCTATTACATAGAGTGTGCTGCTGATGAGATATGACCACTGTGACTGTAGC 120  
QY 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAACGCTTAAACAAACAAATGAAGACTTG 180  
DB 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAACGCTTAAACAAACAAATGAAGACTTG 180  
QY 181 AGGAGATGCGTGTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240  
DB 181 AGGAGATGCGTGTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240  
QY 241 GGAAGATGCGTGTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGCCAGCTTCA 300  
DB 241 GGAAGATGCGTGTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGCCAGCTTCA 300  
QY 301 AAGAAACAACATAAAATGCTCCAGGAGAGATCATGAATCCTTAAGAGAGAGATTGAA 360  
DB 301 AAGAAACAACATAAAATGCTCCAGGAGAGATCATGAATCCTTAAGAGAGAGATTGAA 360  
QY 361 AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTTA 420  
DB 361 AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTTA 420  
QY 421 GAAGAAATGAAGCTCCAAAGACATGAGATGATTTCTATGATTTGGACATTCATGAA 480  
DB 421 GAAGAAATGAAGCTCCAAAGACATGAGATGATTTCTATGATTTGGACATTCATGAA 480  
QY 481 AGGTCTATATGACGATCTATCTACTCAGTCAACAGATGGGCGAGGTGATGGCGT 540  
DB 481 AGGTCTATATGACGATCTATCTACTCAGTCAACAGATGGGCGAGGTGATGGCGT 540

DB 481 AGGTCTATATGACGATCTATCTACTCAGTCAACAGATGGGCGAGGTGATGGCGT 540  
QY 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTGTCACCGGAGATTAACATATCTTCAAGAT 600  
DB 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTGTCACCGGAGATTAACATATCTTCAAGAT 600  
QY 601 CCCAAGAGCTGCGCAAGAGCAAGAGCTAGTGTGTAATATCAACAAAGGCTGTGCTAT 660  
DB 601 CCCAAGAGCTGCGCAAGAGCAAGAGCTAGTGTGTAATATCAACAAAGGCTGTGCTAT 660  
QY 661 GCGTGTACCTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
DB 661 GCGTGTACCTCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
QY 721 CTGCGCTTGAATCTCAATTTGGGCTGACGCTACGCTACGCTACGCTACGCTACGCTACG 780  
DB 721 CTGCGCTTGAATCTCAATTTGGGCTGACGCTACGCTACGCTACGCTACGCTACGCTACG 780  
QY 781 CCTGTAAGTGAAGAGCTGACAGAGATGCTGGCAGCTCCAGCTGACATTTGGTGAAGAA 840  
DB 781 CCTGTAAGTGAAGAGCTGACAGAGATGCTGGCAGCTCCAGCTGACATTTGGTGAAGAA 840  
QY 841 GTAAGAGACAAAATGTTGAGTGTGAGCTTCCCATTTGATAGACAGTGTTCATCCCTCGT 900  
DB 841 GTAAGAGACAAAATGTTGAGTGTGAGCTTCCCATTTGATAGACAGTGTTCATCCCTCGT 900  
QY 901 CCTCATATTTTACCCCTGCTGCTCCAGAGACCTTTCAGATGATGATGATGATGATGATGAT 960  
DB 901 CCTCATATTTTACCCCTGCTGCTCCAGAGACCTTTCAGATGATGATGATGATGATGATGAT 960  
QY 961 GGTGATCTGCACTGT 1020  
DB 961 GGTGATCTGCACTGT 1020  
QY 1021 TGCGTGGAAAGAGAAATAGAGAGGCGCACCAAGAGCTTCAACATCCAGTTATTT 1080  
DB 1021 TGCGTGGAAAGAGAAATAGAGAGGCGCACCAAGAGCTTCAACATCCAGTTATTT 1080  
QY 1081 GGAATCATGTTTGAAGCAGACAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
DB 1081 GGAATCATGTTTGAAGCAGACAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
QY 1141 GAATACAGGCTGACGTTGGAAGAACTTTCAGCTTCTTCTGCGAATGCAAGTGGAT 1200  
DB 1141 GAATACAGGCTGACGTTGGAAGAACTTTCAGCTTCTTCTGCGAATGCAAGTGGAT 1200  
QY 1201 AAAAAAAGGATGATTTTGGCCACAGATGACCTGCTTGTGTAAGAGAGGCAAAACAAAG 1260  
DB 1201 AAAAAAAGGATGATTTTGGCCACAGATGACCTGCTTGTGTAAGAGAGGCAAAACAAAG 1260  
QY 1261 TACCCAGTTATGAAATTTTATGATGATTAATCTATCTCTTGTGACCTGACACTACATTAAT 1320  
DB 1261 TACCCAGTTATGAAATTTTATGATGATTAATCTATCTCTTGTGACCTGACACTACATTAAT 1320  
QY 1321 CGATATACAGAAATTTTACCTGCGGAGTGTGATGATGATGATGATGATGATGATGATGAT 1380  
DB 1321 CGATATACAGAAATTTTACCTGCGGAGTGTGATGATGATGATGATGATGATGATGATGAT 1380  
QY 1381 GACTTCTAGTGTGATCTTTTTCATGCGAGGCTGTGATGATGATGATGATGATGATGATGAT 1440  
DB 1381 GACTTCTAGTGTGATCTTTTTCATGCGAGGCTGTGATGATGATGATGATGATGATGATGAT 1440  
QY 1441 GCGCTCATCTGATGCTCTGCGAAGCTTCCGTTCTTGTGATGACATCTACTATTTTGGGA 1500  
DB 1441 GCGCTCATCTGATGCTCTGCGAAGCTTCCGTTCTTGTGATGACATCTACTATTTTGGGA 1500  
QY 1501 GCGCCAAATGCGCCAAACCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
DB 1501 GCGCCAAATGCGCCAAACCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
QY 1561 ATCCCATGGAACCTGAGATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620  
DB 1561 ATCCCATGGAACCTGAGATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620

QY 1621 AAGGTGTTAACAGAAACTGGGAGAGCGGCTATATCCCTCTACAAAGTTCCAGAG 1680  
|||||  
Db 1621 AAGGTGTTAACAGAAACTGGGAGAGCGGCTATATCCCTCTACAAAGTTCCAGAG 1680  
QY 1681 AAGATGAAACAGTCAAGTACCCCATATATCCCGAGCTGACAGTAA 1728  
|||||  
Db 1681 AAGATGAAACAGTCAAGTACCCCATATATCCCGAGCTGACAGTAA 1728  
RESULT 3  
US-08-913-805A-9  
Sequence 9, Application US/08913805A  
Patent No. 6054304  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: UOZUMI, Naotomi  
APPLICANT: SHIBA, Tetsuo  
APPLICANT: YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetli, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-913-805A-9  
Query Match 87.1%; Score 1505.6; DB 3; Length 2100;  
Best Local Similarity 92.0%; Pred. No. 0;  
Matches 1589; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

|||||  
Db 258 TTGCTCTTTTATATAGGTGCTCACTTGTACGAGATAATGACCATCTGTCCTAGC 317  
QY 121 CGAGACTGTCGAAGATTTTGGCAAGCTGGAAAGCTTAAACAAAGAAATGAAGCTTG 180  
|||||  
Db 318 CGAGACTGTCGAAGATTTTGGCAAGCTTGAAGCTTAAACAGCAAGATGAAGCTTG 377  
QY 181 AGGAGATGAGCTTGATCTCTCCGAATACGAGAGGCCCATTTGATGAGGGCCAGCTTCA 240  
|||||  
Db 378 AGGCAATGAGCCGCAATCTCTCCGGATACCGAAGGCCCTTTTATGACGGCCAGCTATA 437  
QY 241 GGAAGATTGCTGCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTGAAATTTAT 300  
|||||  
Db 438 GGAAGATTGCTGCTTTAGAGAGCAAGCTTTGAAAGGCCCAAGACAGATTGAAATTTAC 497  
QY 301 AAGAACAACCTTAAATATGCTCCAGGAAGGATCATGAATCTTAAGAGAGAGATTGAA 360  
|||||  
Db 498 AAGAACAACCTTAAATATGCTCCAGGAAGGATCATGAATCTTGAAGAGAGATTGAA 557  
QY 361 AATGAGCTTAAAGCTCTGCTTTTCTCAAGTGAGTGAAGAAATTAAGAAATTTA 420  
|||||  
Db 558 AATGAGCTTAAAGCTCTGCTTTTCTCAAGTGAGTGAAGAAATTAAGAAATTTA 617  
QY 421 GAAGAAATGAACTCCAAAGACATGCAGATGAATTTCTATCAGATTGGGACATGAA 480  
|||||  
Db 618 GAAGAAATGAACTCCAAAGACATGCAGATGAATTTCTTGGATTGAGACATCATGAA 677  
QY 481 AGGTCTATATAGAGATCTATCTACCTACCTCAAGCTCAAGATGAGGAGGAGTGGCGT 540  
|||||  
Db 678 AGGTCTATATAGAGATCTATCTACCTACCTCAAGCTCAAGATGAGGAGGAGTGGCGG 737  
QY 541 GAAGAAGAGGCCAAAGATCTGACAGAGCTGCTCAGCGGGAATTAATATCTTCAGAAAT 600  
|||||  
Db 738 GAAGAAGAGGCCAAAGATCTGACAGAGCTGCTCAGCGGGAATTAATATCTTCAGAAAT 797  
QY 601 CCCAAGACTGACAGCAAGCCAAAGCTAGTGTATATCAACAAGGCTGTGCTAT 660  
|||||  
Db 798 CCCAAGACTGACAGCAAGCCAAAGCTAGTGTATATCAACAAGGCTGTGCTAT 857  
QY 661 GGCTGTACGCTCATCATAGTGTAGTACTGCTTATGATTTGATGAGGACCCAGCGACA 720  
|||||  
Db 858 GGCTGTACGCTCATCATAGTGTAGTACTGCTTATGATTTGATGAGGACCCAGCGACA 917  
QY 721 CTCGCTTGGAAATCTCAGAAATTTGGCGTACGCTACTGCGGAGATGGGAAATCTGTTAGA 780  
|||||  
Db 918 CTCATCTTGGAAATCTCAGAAATTTGGCGTATGCTACTGCGGAGATGGGAAATCTGTTAGG 977  
QY 781 CCTGTAAAGTGAAGCTGCACAGACAGATCTGCAAGCTCCACTGACATTTGTCAGGTGAA 840  
|||||  
Db 978 CCTGTAAAGTGAAGCTGCACAGACAGATCTGCAAGCTCCACTGACATTTGTCAGGTGAA 1037  
QY 841 GTAAAGGACAAATATGTCAGTGGTGTGAGCTGCCCATTTAGACAGTGTCAATCTCGT 900  
|||||  
Db 1038 GTAAAGGACAAATATGTCAGTGGTGTGAGCTGCCCATTTAGACAGTGTCAATCTCGT 1097  
QY 901 CCTCATATTTTACCCTGCTGTCGCAAGAGACCTTGCAGATGACATTTGAGAGTCCAT 960  
|||||  
Db 1098 CCTCATATTTTACCCTGCTGTCGCAAGAGACCTTGCAGATGACATTTGAGAGTCCAT 1157  
QY 961 GGTGATCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
|||||  
Db 1158 GGTGATCTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1217  
QY 1021 TTGCTGGAAGAAATATGAGAGGAGGACCAAGAGCTTAGCTTCAAAATCTCAGTTAT 1080  
|||||  
Db 1218 TTGCTGGAAGAAATATGAGAGGAGGACCAAGAGCTTAGCTTCAAAATCTCAGTTAT 1277  
QY 1081 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGCGGAACAGCTTCCATCCATTTGAG 1140  
|||||  
Db 1278 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGCGGAACAGCTTCCATCCATTTGAA 1337  
QY 1141 GAATPACAGGTGACGTTGGAAGAGACTTTCAGCTTCTGCGACAAATGCAAGTGGAT 1200  
|||||



```

Db 1338 GAGTACATGGGTCATGTTGAAGAACATTTTCACGTTCTCTGCACGAGAAATGCA
Qy 1201 AAAAAAAGGCGTATTTTGGCCACAGATGACCCGCTTTGTTAAAGAGCAAA
Db 1398 AAAAAAGAGGTATTTGGCCACAGATGACCCCTTTTATTAAGAGGACAA
Qy 1261 TACCCAGTTATGAATTTATTTAGGATACGTATCTGTTGGGTCAGCTGAGACT
Db 1458 TACCCCAATTATGATTTATTTAGGATACGTATTTCTTGGGTCAGCTGAGACT
Qy 1321 CGATATACAGAAAATTCACCTTCGGGGGTGTGATCCTGGATATACACTTCTCTC
Db 1518 CGATACACAGAAATTTACCTTCGGGAGTATCCTGGATATACATTTTCTCTC
Qy 1381 GACTTCTAGTGTGTACTTTTTCATGCGACAGCTGTATGAGATTGCTTATGAAT
Db 1578 GACTTCTAGTGTGTACTTTTTCATCCCAAGCTGTGTGAGATTGCTTATGAAT
Qy 1441 GCGTGCATCCTGATGGCTCTCGCGAATTCCTGTTCTTTGGATACATCTACTAT
Db 1638 ACATACATCCTGTATGGCTCTCGCAAACTTCATCTTTAGATGACATCTACTAT
Qy 1501 GGCCCAATGGCCCAACCAAAATTCGCAATTTATCTCTACCAACCTGAGCTGAT
Db 1698 GGCCAGATGGCCCAACATTCAAATTTGCCATTTATGCTTCACCAACCCGAACTG
Qy 1561 ATCCCATGGAACCTGAGATATTTATTTGATGTGTGGCTGGAATACATGGATGG
Db 1758 ATCCCATGGAACCTGAGATATCATTTGATGTGTGGCTGGAATATCATTTGGATGG
Qy 1621 AAAGTGTATACAGAAAATCTGGAAGAGCGGCGCTATATCCCTCTACAAAGT
Db 1818 AAAGGTGTCAACAGAAAATTTGGAGAGCGGCGCTATATCCCTCTACAAAGT
Qy 1681 AAGATTAACACAGTCAAGTACCCACATATCCGACGAGCTGACAAAGTAA 1728
Db 1878 AAGTATGAACAGGTCAAGTACCCACATATCTCTGAGGCTGAGAAATTA 1925

RESULT 4
US-09-442-629-9
: Sequence 9, Application US/09442629
: Patent No. 6291219
: GENERAL INFORMATION:
: APPLICANT: TANIUCHI, Naoyuki
: UOZUMI, Naofumi
: SHIBA, Tetsuo
: YAMAGICHI, Shusaku
: TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W., Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: US
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3+ Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
: SOFTWARE: WordPerfect 6.1 Windows
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/442,629
: FILING DATE: 18-NO. 6291219-1999
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/913,805A
: FILING DATE: 7 JAN 1998
: APPLICATION NUMBER: PCT/JP97/00171
: FILING DATE: 23 JAN 1997
: APPLICATION NUMBER: JP 192260
: FILING DATE: 22 JUL 1996
: APPLICATION NUMBER: JP 162813

```

:		FILING DATE: 24 JUN 1996	
:		APPLICATION NUMBER: JP 161648	
:		FILING DATE: 21 JUN 1996	
:		APPLICATION NUMBER: JP 10365	
:		FILING DATE: 24 JAN 1996	
:		ATTORNEY/AGENT INFORMATION:	
:		NAME: Toifenettl, Judith L.	
:		REGISTRATION NUMBER: 39,048	
:		REFERENCE/DOCKET NUMBER: 2356/3	
:		TELECOMMUNICATION INFORMATION:	
:		TELEPHONE: 202-429-1776	
:		TELEFAX: 202-429-0796	
:		INFORMATION FOR SEQ ID NO: 9:	
:		SEQUENCE CHARACTERISTICS:	
:		LENGTH: 2100 base pairs	
:		TYPE: nucleic acid	
:		STRANDEDNESS: double	
:		TOPOLOGY: linear	
:		MOLECULE TYPE: cDNA	
:		SEQUENCE DESCRIPTION: SEQ ID NO: 9:	
:		US-09-442-629-9	
:			
:		Query Match	E7.1%: Score 1505.6; DB 4; Length 2100;
:		Best Local Similarity	52.0%; Pred No. 0; Mismatches 139; Indels 0; Gaps 0;
:		Matches 1589; Conservative	0;
:QY	1	ATGCGGCATGAGACGTGGTTCGCGCTTGCGATATGCTCATTTCTTTTGGCTGGGGACC	60
Db	198	ATGCGGCATGAGACGTGGTTCGCGCTTGCGATATGCTCATTTCTTTTGGCTGGGGACC	257
:QY	61	TTGCTATTTTACATAGGTGGTCACTTGGTAGCAGATAATGACACTGTGACTCTTAC	120
Db	258	TTGCTATTTTATATAGGTGGTCACTTGGTAGCAGATAATGACACTGTGACTCTTAC	317
:QY	121	CGAGAACGTCCAAATTTTGGCAAGCGGAAGCCTTAACAACAAATAAAGACTTG	180
Db	318	CGAGAACGTCCAAATTTTGGCAAGCGCTTAACAACACAGATTAAGAAGCTTG	377
:QY	181	AGSAGAAATGGCGTGAATCTCCGATTAACAGAAAGGCCCATTTGATCAGGGCCAGCTTCA	240
Db	378	AGGCGAATGGCGGATCTCTCCGATTAACAGAAAGGCCCATTTGATCAGGGCCAGCTTAA	437
:QY	241	GGAAGAGTTTCGTCTTGAAGAGAGCAATTTATGAAGCCCCAAAGAACAGATTGAAAAATTAT	300
Db	438	GGAAGAGTACCGCTTTTGAAGAGAGCACTGTTAAGGCCAAAGAACAGATTGAAAAATTAC	497
:QY	301	AAGAAACAACATTAATAATGGTCCAGGAAAGATCTATGAATTCCTAAGAGAGAGATTGAA	360
Db	498	AAGAAACAAGACCGAAATATGGTCTGGGGAAGATCATGAATCTCTAAGAGAGAGATTGAA	557
:QY	361	AATGAGCTAAAGACCTCTGGTTTTTCTACAAAGAGATTAAGAAATTAAGAATTAA	420
Db	558	AATGAGCTAAAGACCTCTGGTTTTTCTACAGAGATTAAGAAATTAAGAATTAA	617
:QY	421	GAAGAAATGAACCTCCAAGACATGACAGATGAATTTCTATCAGATTTGGGACATCATGAA	480
Db	618	GAAGAAATGAACCTCCAAGACATGACAGATGAATTTCTTTCGATTTAGGACATCATGAA	677
:QY	481	AGGTTATTAATGACGATCTATCTACTCACTCACTCAACAGATGGGGCAGGTATGGCGT	540
Db	678	AGGTTATTAATGACGATCTATCTACTCACTCACTCAACAGATGGGGCAGGTATGGCGG	737
:QY	541	GAAGAAGAGGCCAAAGATCTGACAGAGCGTCCAGCGGAGAAATTAACATATCTTCAGAT	600
Db	738	GAAGAAGAGGCCAAAGATCTGACAGAGCGTCCAGCGGAGAAATTAACATATCTTCAGAT	797
:QY	601	CCCAAAGGACTGACAGAAAGCCAAAGAGCTAGTGTGTAATATCAACAAAAGGCTGTGGCTAT	660
Db	798	CCCAAAGGACTGACAGAAAGCCAAAGAGCGTGTGTAATATCAACAAAAGGCTGTGGCTAT	857
:QY	661	GGCTGTCAAGCTCATCATGTAGTACTGATCTGATTTATGATTTGCATTATGGCACCCAGCGAACA	720
Db	858	GGCTGTCAAGCTCATCATGTAGTACTGATCTGATTTATGATTTGCATTATGGCACCCAGCGAACA	917



QY 587 CATATCTCAGATCCCAAGSACTGAGCAAGCAAGAGTGTGTAA 638  
DB 1070 RRRATCGCAAGCTCCCTCGACTGAGCAAGCTCGAATTATCTGTGA 1019

## RESULT 6

US-08-559-896B-3

Sequence 3, Application US/08559896B

Patent No. 6310046

GENERAL INFORMATION:

APPLICANT: Patrick E. Duffy

APPLICANT: Christian F. Ockenhouse

TITLE OF INVENTION: SEQUESTIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: John Moran

STREET: USA MRC - MCMR-JA

CITY: FORT DETRICK, FREDERICK

STATE: MARYLAND

COUNTRY: USA

ZIP: 21702-5012

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.5

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/559, 896B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Moran, John

REGISTRATION NUMBER: 26,313

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 619-7714

TELEFAX: (301) 619-7714

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 417 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

US-08-559-896B-3

Query Match 2.2%; Score 38.2; DB 4; Length 417;  
Best Local Similarity 47.7%; Pred. No. 0.036; Indels 0; Gaps 0;  
Matches 112; Conservative 0; Mismatches 123;

QY 258 AGAAGCAATTTATGAGGCCCAAGACAGATTGAATTTATAGAACAACTAATAA 317  
DB 183 AGAAGCAATTTATGAGGCCCAAGACAGATTGAATTTATAGAACAACTAATAA 242  
QY 318 TGGTCAGGGAAGGATCATGAATCTCTAAGAGAGGAGATTGAAGCTAAGAGCT 377  
DB 243 TAAACTTCAAGAGCAAGCAATTAATTAATGATGATCAAGCTGATTAAGATAG 302  
QY 378 CTGGTTTTTCTACAAAGTGAGTTGAAGAAATTAAGAAATTTAGAGAAATGAATCCA 437  
DB 303 AGAAGCAATTTAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 362  
QY 438 AAGACATGAGATGATTTCTATCAGATTTGGGACATCATGAAGCTCTATATG 492  
DB 363 TAAATATATAGAGAGATTTAGACAAATGATGATGCAATTTATATAGATG 417

RESULT 7  
US-08-559-896B-1  
Sequence 1, Application US/08559896B

Patent No. 6310046  
GENERAL INFORMATION:  
APPLICANT: Patrick E. Duffy  
APPLICANT: Christian F. Ockenhouse  
TITLE OF INVENTION: SEQUESTIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John Moran  
STREET: USA MRC - MCMR-JA  
CITY: FORT DETRICK, FREDERICK  
STATE: MARYLAND  
COUNTRY: USA  
ZIP: 21702-5012

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.5

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/559, 896B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Moran, John

REGISTRATION NUMBER: 26,313

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 619-7714

TELEFAX: (301) 619-7714

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1956 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

US-08-559-896B-1

Query Match 2.2%; Score 38.2; DB 4; Length 1956;  
Best Local Similarity 47.7%; Pred. No. 0.095;  
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 258 AGAAGCAATTTATGAGGCCCAAGACAGATTGAATTTATAGAACAACTAATAA 317  
DB 306 AGAAGCAATTTATGAGGCCCAAGACAGATTGAATTTATAGAACAACTAATAA 365  
QY 318 TGGTCAGGGAAGGATCATGAATCTCTAAGAGAGGAGATTGAAGCTAAGAGCT 377  
DB 366 TAAACTTCAAGAGCAAGCAATTAATTAATGATGATGATCAAGCTGATTAAGATAG 425  
QY 378 CTGGTTTTTCTACAAAGTGAGTTGAAGAAATTAAGAAATTTAGAGAAATGAATCCA 437  
DB 426 AGAAGCAATTTAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 485  
QY 438 AAGACATGAGATGATTTCTATCAGATTTGGGACATCATGAAGCTCTATATG 492  
DB 486 TAAATATATAGAGAGATTTAGACAAATGATGATGCAATTTATATAGATG 540

RESULT 8  
US-08-037-816A-19  
Sequence 19, Application US/08037816A  
Patent No. 5869624.  
GENERAL INFORMATION:  
APPLICANT: Hasel, Karl W.  
APPLICANT: Madden, Paul J.  
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:



APPLICATION NUMBER: US/08/037,816  
FILING DATE: 26-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41190/JPM/AJM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOPUI  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1448 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1439  
OTHER INFORMATION:  
US-08-530-146-19

Query Match 2.2%; Score 37.8; DB 2; Length 1448;  
Best Local Similarity 52.9%; Pred. No. 0.11;  
Matches 81: Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTTATGAAGGCCAAGACAGATTGAAATTTAAGAAACAACATA 313  
DB 944 CTTTAAACAGATAGTTATTAATTAAGAGCAATTTGAGATTAACAAATGCTTTA 1003

QY 314 AAAATGCTCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAAATGGAGCTAAG 373  
DB 1004 ATCAGCTCCAGAGAGGAGCCAGAAATGTATGCGACAGTTTAATTTGGAGAGAAAT 1063

QY 374 AGCTGTGTTTTCTACAAAGTAGTTGAGA 406  
DB 1064 TTTTCTACTGTATTCACACAACTGTTTAATA 1096

RESULT 11  
US-08-530-146-23  
Sequence 23, Application US/08530146  
Patent No. 5886163  
GENERAL INFORMATION:  
APPLICANT: Hasel, Karl W.  
APPLICANT: Maddon, Paul J.  
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,146  
FILING DATE: 26-MAR-1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/037,816  
FILING DATE: 26-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41190/JPM/AJM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOPUI  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1448 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1438  
OTHER INFORMATION:  
US-08-530-146-23

Query Match 2.2%; Score 37.8; DB 2; Length 1448;  
Best Local Similarity 52.9%; Pred. No. 0.11;  
Matches 81: Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTTATGAAGGCCAAGACAGATTGAAATTTAAGAAACAACATA 313  
DB 944 CTTTAAACAGATAGTTATTAATTAAGAGCAATTTGAGATTAACAAATGCTTTA 1003

QY 314 AAAATGCTCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAAATGGAGCTAAG 373  
DB 1004 ATCAGCTCCAGAGAGGAGCCAGAAATGTATGCGACAGTTTAATTTGGAGAGAAAT 1063

QY 374 AGCTGTGTTTTCTACAAAGTAGTTGAGA 406  
DB 1064 TTTTCTACTGTATTCACACAACTGTTTAATA 1096

RESULT 12  
US-08-037-816A-15  
Sequence 15, Application US/08037816A  
Patent No. 5869624  
GENERAL INFORMATION:  
APPLICANT: Hasel, Karl W.  
APPLICANT: Maddon, Paul J.  
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/037,816A  
FILING DATE: 26-MAR-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41190/JPM/AJM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOPUI  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1532 base pairs

```

:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 1..1522
:      OTHER INFORMATION:
US-08-037-816A-15

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTATGAGCCAAAGACAGATTGAAATTTAAGAAACAACCTA 313
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 CTTTAAACAGATAGTATTAATAATTAAAGACAACTTTGAGATTAACAACTGCTTTA 1087

QY 314 AAATGGTCCAGGAGGAGATGAAATCCTTAAGAGAGAGATTGAAATGAGCTAAAG 373
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1088 ATCAGCTCCTCAGGAGGAGACCAGAAATGTATGACAGAGTTTAAATGTGAGAGAGAAAT 1147

QY 374 AGCTCTGCTTTTCTTACAAAGTAGTGAAGA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1148 TTTTCTAGCTATATTCAACACACTGTTTAATA 1180

RESULT 13
US-08-037-816A-27
: Sequence 27, Application US/08037816A
: Patent No. 5869624
: GENERAL INFORMATION:
: APPLICANT: Hasel, Karl W.
: TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
: TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/037,816A
: FILING DATE: 26-MAR-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOPUI
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1532 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1522
: OTHER INFORMATION:

```

```

US-08-037-816A-27

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTATGAGCCAAAGACAGATTGAAATTTAAGAAACAACCTA 313
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 CTTTAAACAGATAGTATTAATAATTAAAGACAACTTTGAGATTAACAACTGCTTTA 1087

QY 314 AAATGGTCCAGGAGGAGATGAAATCCTTAAGAGAGAGATTGAAATGAGCTAAAG 373
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1088 ATCAGCTCCTCAGGAGGAGACCAGAAATGTATGACAGAGTTTAAATGTGAGAGAGAAAT 1147

QY 374 AGCTCTGCTTTTCTTACAAAGTAGTGAAGA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1148 TTTTCTAGCTATATTCAACACACTGTTTAATA 1180

RESULT 14
US-08-530-146-15
: Sequence 15, Application US/08530146
: Patent No. 5886163
: GENERAL INFORMATION:
: APPLICANT: Hasel, Karl W.
: TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
: TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/530,146
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/037,816
: FILING DATE: 26-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOPUI
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1532 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1522
: OTHER INFORMATION:
US-08-530-146-15

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```

Oy	254	CTTTAGGAAGGCAATTTATTAAGGCCAAAGACAGATTGAAAAATTATGAGAAACAACCTA	313
Db	1028	CTTTAAACACAGATGATTATTAATTAAGAGAACAAATTTAAGAAATTAACAAATAGCTCTTTA	1087
Oy	314	AAATGGTCACAGGGAAGCATCATGAATCTTAAGAGAGAGATTGAAATGGACCTTAAG	373
Db	1088	ATCAGCTCCTCAGGAGGGAGACCCAGAAATTTGTATATGCACAGTTTAAATTTGTGGAGAGAAAT	1147
Oy	374	AGCTCTGGTTTTCTTCACAAATGATGTGACAG	406
Db	1148	TTTTCTACTGAATTCACACAACTGTTTATATA	1180

```

Oy  314  AAAATGGCCAGGGAAGATCATGAAATCCTAAGAGGAGGATTTGAAAATGGAGCTAAAG  373
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  1088  ATCACTCCTCAGGAGGGAGACCAGAAATTTGATTCACAGTTTTTAATTTGGAGCAGANT  1147
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy  374  AGCTCTGGTTTTTCTACAAAGTGAAGTTGAGA  406
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  1148  TTTTCTACTGTATTCACACACAACTGTTTATA  1180
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Search completed: November 1, 2002, 23:06:58  
Job time : 93 secs

```
Search completed: November 1, 2002, 23:06:58
Job time : 93 secs
```

RESULT 15  
 US-08-530-146-27  
 Sequence 27, Application US/08530146  
 Patent No. 5886163  
 GENERAL INFORMATION:  
 APPLICANT: Hasel, Karl W.  
 APPLICANT: Maddou, Paul J.  
 TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED  
 TITLE OF INVENTION: THERIO, AND THERAPEUTIC AND PROPHYLACTIC USES  
 TITLE OF INVENTION: THEREOF  
 NUMBER OF SEQUENCES: 29  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Cooper & Dunham  
 STREET: 30 Rockefeller Plaza  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10112  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.24  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/530,146  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/037,816  
 FILING DATE: 26-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 41190/JPW/AJM  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 977-9550  
 TELEFAX: (212) 664-0525  
 TELEX: 422523 COOPUI  
 INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1532 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1522  
 OTHER INFORMATION:  
 US-08-530-146-27

Query Match	2.2%	Score 37.8;	DB 2;	Length 1532;
Best Local Similarity	52.9%	Pred. NO. 0.11;		
Matches	81; Conservative	0;	Mismatches 72;	Indels 0; Gaps 0;

Qy	254	CTTTAGAGACCAATTTTATGAGGCCAAGCAACAGATTGMAATTTATTAAGAAACAACCTA	313
Db	1028	CTTTAAACGATAGTTATATAATTAAGAGAACAATTTGAGATTAACAACATAGCTTTTA	1087

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 20:05:17 : Search time 264 Seconds

(without alignments)  
14740.353 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728

Sequence: 1 atgcgcgcactgactgctc.....atccgcagcgcacagtaaa 1728

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726.4	99.9	1728	18	AA76573
2	1510.4	87.4	1759	24	ABA98809
3	1510.4	87.4	2100	18	AA76574
4	1510.4	87.4	3291	23	ABV22468
5	1510.4	87.4	3291	23	ABV28283
6	1499	86.7	3007	22	AAH3315
7	1472.6	85.2	2008	24	ABK70031
8	1467.8	84.9	1728	24	ABK70032
9	878.4	50.8	1017	22	AA87952

10	833.4	48.2	979	21	AA63891
11	833.4	48.2	979	24	ABK70036
12	827	47.9	979	21	AA63892
13	827	47.9	979	24	ABK70037
14	606.8	35.1	699	22	AA87953
15	351.6	20.3	2761	22	ABL04601
16	288.4	16.7	394	20	AAV89049
17	275.2	15.9	503	22	AAK11265
18	275.2	15.9	503	22	AAI42879
19	218.8	12.7	551	22	AAK11037
20	218.8	12.7	551	22	AAI42659
21	215	12.4	4682	23	ABL04600
22	213.4	12.3	248	22	ABA47883
23	213.4	12.3	248	22	ABA65776
24	213.4	12.3	248	22	ABA32861
25	213.4	12.3	248	22	AAK14177
26	213.4	12.3	248	22	AAK25142
27	213.4	12.3	248	22	AAK39911
28	213.4	12.3	248	22	AAI20722
29	213.4	12.3	248	22	AAI45939
30	213.4	12.3	248	22	AAI57183
31	213.4	12.3	248	22	AAI06420
32	213.4	12.3	248	22	ABSI4010
33	174.2	10.1	9196	24	ABK70033
34	139.6	8.1	384	22	ABA42755
35	139.6	8.1	384	22	ABA53183
36	139.6	8.1	384	22	ABA22955
37	139.6	8.1	384	22	AAK01429
38	139.6	8.1	384	22	AAK26877
39	139.6	8.1	384	22	AAI11508
40	139.6	8.1	384	22	AAI33785
41	139.6	8.1	384	22	AAI01422
42	139.6	8.1	384	24	ABS01477
43	115.8	6.7	129	22	AAK23927
44	115.8	6.7	129	22	AAI55934
45	58.4	3.4	60	24	ABN32244

#### ALIGNMENTS

RESULT 1  
AAT76573  
AAT76573 standard; cDNA to mRNA; 1728 BP.

05-MAR-1998 (first entry)

Pig alpha 1-6 fucosyltransferase gene.

Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
guanosine diphosphate; sugar chain synthesis; modification; antibody;

GlcNAc; cancer diagnosis; ss.

Sus scrofa.

Key Location/Qualifiers  
CDS 1..1728  
/\*tag= a

W09727303-A1.

31-JUL-1997.

23-JAN-1997; 97WO-JP00171.

22-JUL-1996; 96JP-0193260.

24-JAN-1996; 96JP-0010365.

21-JUN-1996; 96JP-0161648.

24-JUN-1996; 96JP-0162813.

(TOYM ) TOYO BOSEKI KK.

Chinese hamster FU  
Antibody productio  
Rat F078 cDNA. Ra  
Antibody productio  
Human alpha 1-6 fu  
Drosophila melunog  
EST clone CH100.  
Human brain expres  
Probe #11565 used  
Human brain expres  
Human brain expres  
Drosophila melano  
Human breast cell  
Human foetal liver  
Probe #11327 for g  
Human brain expres  
Human brain expres  
Human bone marrow  
Probe #10655 for g  
Probe #14625 used  
Probe #25869 used  
Probe #6411 used t  
Human genome-deriv  
Antibody productio  
Human breast cell  
Human foetal liver  
Probe #1421 for ge  
Human brain expres  
Human bone marrow  
Probe #1441 for ge  
Probe #1473 used t  
Human genome-deriv  
Human brain expres  
Probe #24620 used  
Human spliced tran

XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
PI WPI: 1997-393690/36.  
DR P-PSDB; AAM22124.  
XX  
XX Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for  
PT synthesis and modification of sugar chains and used as an antigen  
PT for production of diagnostic antibodies  
XX  
PS Claim 5: Page 30-34; 61pp: Japanese.  
XX  
XX AAT76573 and AAT76574 represent the coding sequences for the pig and  
CC human alpha 1-6 fucosyltransferases of the invention, respectively. The  
CC encoded enzyme transfers fucose from guanosine diphosphate to the  
CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:  
CC (GlcNAc-beta 1-2-Man-alpha 1-6)(GlcNAc-beta 1-2-Man-alpha 1-3)Man-beta  
CC 1-4(GlcNAc-beta 1-4(GlcNAc-R to give (GlcNAc-beta 1-2-Man-alpha 1-6)(GlcNAc-beta  
CC 1-2-Man-alpha 1-3)Man-beta 1-4(GlcNAc-beta 1-4(Fucose-alpha 1-6)GlcNAc-R. It has  
CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the  
CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working  
CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not  
CC required for activity of the enzyme, and the enzyme is not inhibited in  
CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
CC modification of sugar chains, and as antigen for the production of  
CC antibodies recognising the enzyme. The antibodies can be used for the  
CC diagnosis of cancer and other diseases.  
XX  
XX Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;

Query Match 99.9%; Score 1726.4; DB 18; Length 1728;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGGGCCATGAGCTGGTGGCGTGGATATGCTCATCTCTTTTGGCTGGGGACC 60  
DB 1 ATCGGGCCATGAGCTGGTGGCGTGGATATGCTCATCTCTTTTGGCTGGGGACC 60  
QY 61 TTGCTATTTTACATAGTGGTGCATCTGGTACGAGATATAGCCACTGTCACTAGC 120  
DB 61 TTGCTATTTTACATAGTGGTGCATCTGGTACGAGATATAGCCACTGTCACTAGC 120  
QY 121 CGAGAACTGTCAGATTTTGGCAAGCTGGAGCCTTAAACAAATGAAGACTTG 180  
DB 121 CGAGAACTGTCAGATTTTGGCAAGCTGGAGCCTTAAACAAATGAAGACTTG 180  
QY 181 AGAGAGATGGCTGATCTCCGAAATACAGAAAGGCCCTTGTATCAGGGGCGAGCTTCA 240  
DB 181 AGAGAGATGGCTGATCTCCGAAATACAGAAAGGCCCTTGTATCAGGGGCGAGCTTCA 240  
QY 241 GGAAGAGTTCGCTTTAGAAAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 300  
DB 241 GGAAGAGTTCGCTTTAGAAAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 300  
QY 301 AAGAAACAACTAAAAATGTCAGAGGAGATCATGAATCTCTAAGGAGAGATTGAA 360  
DB 301 AAGAAACAACTAAAAATGTCAGAGGAGATCATGAATCTCTAAGGAGAGATTGAA 360  
QY 361 AATGAGACTAAGAGCTCTGTTTTTTCTCAAAAGTCAAGTGAAGAAATTAAGAAATTTA 420  
DB 361 AATGAGACTAAGAGCTCTGTTTTTTCTCAAAAGTCAAGTGAAGAAATTAAGAAATTTA 420  
QY 421 GAAGGAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTGGACATCATGAA 480  
DB 421 GAAGGAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTGGACATCATGAA 480  
QY 481 AGGTCTAATAATGACGATCTATCTACTACCTCAGTCAAAACAGATGGGCGAGTGTGCGGT 540  
DB 481 AGGTCTAATAATGACGATCTATCTACTACCTCAGTCAAAACAGATGGGCGAGTGTGCGGT 540  
QY 541 GAAAGAGAGGCCAAAGATCTGACAGAGCTGTCACAGCGGAGATTAACATATCTTTCAGAAAT 600  
DB 541 GAAAGAGAGGCCAAAGATCTGACAGAGCTGTCACAGCGGAGATTAACATATCTTTCAGAAAT 600

QY 601 CCCAAGACTGCAGCAAAAGCCAAAGACTAGTGTGAATATCAACAAGGCTGTGCTAT 660  
DB 601 CCCAAGACTGCAGCAAAAGCCAAAGACTAGTGTGAATATCAACAAGGCTGTGCTAT 660  
QY 661 GGCTGTACGCTCATCATGTAGTGTACTGCTTTATGATTTGCATATGCGACCCAGCAACA 720  
DB 661 GGCTGTACGCTCATCATGTAGTGTACTGCTTTATGATTTGCATATGCGACCCAGCAACA 720  
QY 721 CTGCGCTTGGAAATCTCAGCAATTTGGCGGTACGCTACTGTGGGATGGGAAACTGTGTTAGA 780  
DB 721 CTGCGCTTGGAAATCTCAGCAATTTGGCGGTACGCTACTGTGGGATGGGAAACTGTGTTAGA 780  
QY 781 CCGTGAAGTGAAGCTGCACAGACAGATCTGCGAGCTTCCACTGGAGATTTGGTCAAGTGA 840  
DB 781 CCGTGAAGTGAAGCTGCACAGACAGATCTGCGAGCTTCCACTGGAGATTTGGTCAAGTGA 840  
QY 841 GTAAGGACAAAAATGTTCAAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900  
DB 841 GTAAGGACAAAAATGTTCAAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900  
QY 901 CCGTCATATTTACCCCGCTGGCTGCCAGAAAGACTTGCAGATGACATGAGTGTGAT 960  
DB 901 CCGTCATATTTACCCCGCTGGCTGCCAGAAAGACTTGCAGATGACATGAGTGTGAT 960  
QY 961 GGTGATCTGTCAGTGTGTGGGTATCCAGTGTGTCAAGTACTTGTGATTCGCCCAACCC 1020  
DB 961 GGTGATCTGTCAGTGTGTGGGTATCCAGTGTGTCAAGTACTTGTGATTCGCCCAACCC 1020  
QY 1021 TGGCTGGAAGAAAGAAATAGAAAGAGGCCACCAAGCTAGGCTTCAAACTCCAGTTAT 1080  
DB 1021 TGGCTGGAAGAAAGAAATAGAAAGAGGCCACCAAGCTAGGCTTCAAACTCCAGTTAT 1080  
QY 1081 GAGATCCATGTTAGAGCAGACAGCAAAAGTGGAGCGGAACAGCTTCCATCCATTTGAG 1140  
DB 1081 GAGATCCATGTTAGAGCAGACAGCAAAAGTGGAGCGGAACAGCTTCCATCCATTTGAG 1140  
QY 1141 GAATACACGCTGCACGTTGAAGAAAGCTTTCAGCTTCTTGCCTGCAGAAATGCAAGTGA 1200  
DB 1141 GAATACACGCTGCACGTTGAAGAAAGCTTTCAGCTTCTTGCCTGCAGAAATGCAAGTGA 1200  
QY 1201 AAAAAAGGCTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAAG 1260  
DB 1201 AAAAAAGGCTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAAG 1260  
QY 1261 TACCCAGTATGAATTTATTTAGTGAATAACTATCTCTTGGTGCAGCTGACTACATTAAT 1320  
DB 1261 TACCCAGTATGAATTTATTTAGTGAATAACTATCTCTTGGTGCAGCTGACTACATTAAT 1320  
QY 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGGATATACACTTTCCTCCAGGCA 1380  
DB 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGGATATACACTTTCCTCCAGGCA 1380  
QY 1381 GACTTCTACTGTGTACTTTTTCATGCGAGCTGTGTAGAGTGTGCTTATGAAATCATGCAA 1440  
DB 1381 GACTTCTACTGTGTACTTTTTCATGCGAGTGTGTAGAGTGTGCTTATGAAATCATGCAA 1440  
QY 1441 GCGCTCATCTGATGCTGTGCGAACTTCCGTTCTTTGGATGACATCTATATTTTGA 1500  
DB 1441 GCGCTCATCTGATGCTGTGCGAACTTCCGTTCTTTGGATGACATCTATATTTTGA 1500  
QY 1501 GGGCCAAATGCCCAACCAAAATTTGCCATTATCTCACCACCTCGAACTGAAGAGAA 1560  
DB 1501 GGGCCAAATGCCCAACCAAAATTTGCCATTATCTCACCACCTCGAACTGAAGAGAA 1560  
QY 1561 ATCCCATGAGACCTGAGATATTTATTTGGTGTGCTGGAATACATGCGATGCTATCT 1620  
DB 1561 ATCCCATGAGACCTGAGATATTTATTTGGTGTGCTGGAATACATGCGATGCTATCT 1620  
QY 1621 AAAGGTGTTAAGCAAAACCTGGGAAGAGCGGCTATATCCCTCTACAAAGTTTCAGAG 1680  
DB 1621 AAAGGTGTTAAGCAAAACCTGGGAAGAGCGGCTATATCCCTCTACAAAGTTTCAGAG 1680



Dd	1277	TACCCCAATTATGAAATTTTATTAGTGATAACTCTAATTTTCCTGGTCAGCTGCACATGCAANT	1336
Oy	1331	CGATATATACAGAAAATTTACTTGGGGGTGTGATCCTGGATATATACACTTCTCTCCAGGCA	1380
Dd	1337	CGATATACAGAAAAATTTACTTCTGTGGAGGTATCTGGATATACATTTTCTCTCACAGCA	1396
Oy	1381	GACTTCCTAGTGTCTACTTTTTTATGTCAGAGTCTGTAGAAGTTGCTTATGAATCATGCAA	1440
Dd	1397	GACTTCCTAGTGTCTACTTTTTTATGTCAGAGTCTGTAGAAGTTGCTTATGAATCATGCAA	1456
Oy	1441	GCCTGCATCCTGATGCCCTCTGCGAACCTTCCGTTCTTTGGATGACATCTACTATTTTGG	1500
Dd	1457	ACACTACATCTCTGATGGCTCTGCGAACCTTCCATCTTTTATGATGACATCTACTATTTTGG	1516
Oy	1501	GGCCCAATATGGCCCCACACAATTTGGCATTATTCCTCACCAACCTGCAAGTGAAGAGCAA	1560
Dd	1517	GGCCCAAGANTGCCCACTACATTAATTTGCCATTTTATGCTCACCAACCCTGCAAGTGAAG	1576
Oy	1561	ATTCCTCATGGAACCTGGAGATATTTATTTGTTGTTGGCTGTGAAGTAATCAGTGGATGGCTATCCT	1620
Dd	1577	ATTCCTCATGGAACCTGGAGATATTTATTTGTTGTTGGCTGTGAAGTAATCATTGGAGGCTATTTCT	1636
Oy	1621	AAAGGTGTTAACAGAAAATCTGGGAGAGGCGCTATATCCTCCTACAAAGTTGGAAG	1680
Dd	1637	AAAGGTGTCAACAGAAAATTTGGGAGAGGCGCTATATCCTCCTACAAAGTTGGAAG	1696
Oy	1681	AAGATTAAGAACAGTCAAGTACCCACATATCCCGAGGCTGCACAACTAA	1728
Dd	1697	AAGATTAAGAACAGTCAAGTACCCACATATCCCGAGGCTGCACAACTAA	1744
<hr/>			
RESULT_3			
AAAT76574			
ID	AAAT76574	standard; cDNA to mRNA; 2100 BP.	
XX	AAAT76574:		
AC	AAAT76574:		
XX			
DT	05-MAR-1998	(first entry)	
XX			
DE	Human alpha 1-6 fucosyltransferase gene.		
XX			
KW	Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;		
KM	guanosine diphosphate; sugar chain synthesis; modification; antibody;		
KX	GlcNAc; cancer diagnosis; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	198..1925	
FT		/tag= a	
XX			
PN	MO9727303-A1.		
PD	31-JUL-1997.		
XX			
PE	23-JAN-1997;	97MO-JP00171.	
XX			
PR	22-JUL-1996;	96JP-0192260.	
PR	24-JAN-1996;	96JP-0010365.	
PR	21-JUN-1996;	96JP-0161448.	
PR	24-JUN-1996;	96JP-0162813.	
XX			
PA	(TOYM ) TOYO BOSEKI KK.		
XX			
PI	Shiba T, Taniguchi N, Uozumi N, Yamagidani S;		
XX			
DR	WPI: 1997-393690/36.		
DR	P-PSDB: AAM22125.		
PT	Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for		
PT	synthesis and modification of sugar chains and used as an antigen		
PT	for production of diagnostic antibodies		
XX*			

PS	Claim 18, Page 39-43; 61pp; Japanese.
XX	
CC	AAT76573 and AAT76574 represent the coding sequences for the pig and
CC	human alpha 1-6 fucosyltransferases of the invention, respectively. The
CC	encoded enzyme transfers fucose from guanosine diphosphate to the
CC	6-hydroxyl group of the GlcNAc nearest to 2-R in the receptor molecule:
CC	(GlcNAc-beta 1-2mannalpha 1-6)(GlcNAc-beta 1-2manalpha 1-3)manbeta
CC	1-4GlcNAc-beta 1-4GlcNAc-R to give (GlcNAc-beta 1-2mannalpha 1-6)(GlcNAc-beta
CC	1-2mannalpha 1-3)manbeta 1-4GlcNAc-beta 1-4(Fuca)alpha 1-6)(GlcNAc-R. It has
CC	an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the
CC	pH range 4-10 after 5 hours at 4 degrees C. The optimum working
CC	temperature of the enzyme is 30-37 degrees C. A bivalent metal is not
CC	required for activity of the enzyme, and the enzyme is not inhibited in
CC	the presence of 5 mM EDTA. The enzyme, and the enzyme is useful in the synthesis and
CC	modification of sugar chains, and as antigen for the production of
CC	antibodies recognising the enzyme. The antibodies can be used for the
CC	diagnosis of cancer and other diseases.
XX	
Sequence 2100 BP; 651 A; 436 C; 489 G; 524 T; 0 other;	
Query Match	87.4%; Score 1510.4; DB 18; Length 2100;
Best Local Similarity	92.1%; Pred. No. 0;
Matches 1592; Conservative	0; Mismatches 136; Indels 0; Gaps 0
OY	1 ATGCGGCATGAGACTGGTTCTGTCGGCTGGATATGCTCATTTCTTTTGGCGGGGACC 60
DB	198 ATGCGGCATGAGACTGGTTCTGTCGGCTGGATATGCTCATTTCTTTTGGCGGGGACC 257
OY	61 TTGCTATTTTACATAGGTGTGCTACTTTGGTACAGATAAAGACCCTGTGATCTCTAC 120
DB	258 TTGCTATTTTAAATAGGTGTGCTACTTTGGTACAGATAAAGACCCTGTGATCTCTAC 317
OY	121 CGAGAACGTGTCCAAATTTTGGCAAGCTGGAACCTTTAAACAACAATGAAGACTTG 180
DB	318 CGAGAACGTGTCCAAATTTTGGCAAGCTGGAACCTTTAAACAACAATGAAGACTTG 377
OY	181 AGGAGATGGCTGTGATCTCTCCGAATPACCAGAAGGCCCATTTGATCAGGGCCACTTCA 240
DB	378 AGGCAATGGCCGAATCTCTCCGATPACCAGAAGGCCCATTTGATCAGGGCCACTTCA 437
OY	241 GGAAGATGTCGCTTTGGAAGAGCAATTTAAGAGGCCCAAGAACAGATTGAAATTTAT 300
DB	438 GGAAGATGTCGCTTTGGAAGAGCAATTTAAGAGGCCCAAGAACAGATTGAAATTTAT 497
OY	301 AAGAAACAACACTAAATAATGTGTCAGAGGAAGATCATGAATCTCTAAGAGAGAGATTGAA 360
DB	498 AAGAAACAACACTAAATAATGTGTCAGAGGAAGATCATGAATCTCTGAGAGAGATTGAA 557
OY	361 AATGAGCTAAAGAGACTCTGGTTTTTCTACAAAGTAGAGTTGAAGAAATTTAAAGATTTA 420
DB	558 AATGAGCTAAAGAGACTCTGGTTTTTCTACAGACTGAATTTGAAGAAATTTAAAGAACTTA 617
OY	421 GAAGGAATTAAGACTCCAAAGACATCAGATGAATTTCTATCGATTTTGGGACATCATGAA 480
DB	618 GAAGGAATTAAGACTCCAAAGACATCAGATGAATTTCTTGGATTTAGGACATCATGAA 677
OY	481 AGGCTATTAATGACGGATCTATATTACTCACTCACTCAACACAGATGGGCGAGTATGGCGT 540
DB	678 AGGCTATTAATGACGGATCTATATTACTCACTCACTCAACACAGATGGGCGAGTATGGCGG 737
OY	541 GAAAGAGGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTTCAGAT 600
DB	738 GAAAGAGGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTTCAGAT 797
OY	601 CCCAAGGACTGAGCAAGGCCAAGAGACTAGTGTATATATCAACAAAGGCTGTGGCTAT 660
DB	798 CCCAAGGACTGAGCAAGGCCAAGAGGCTGTGTATATATCAACAAAGGCTGTGGCTAT 857
OY	661 GGCTGTCAAGCTTCATCATGTAGTGTACTGTCTTTATGATTTGGCATATGGCCAGCGACACA 720
DB	858 GGCTGTCAAGCTTCATCATGTGTGTCTACTGTCTTCAATGATTTGGCATATGGCCAGCGACACA 917
OY	721 CTCGGCTTGAATTCACAAATTTGGGGCTACGCTACTGGGGGATGGGAAACTGTGTTTACA 780



Qy	241	GGAGAGCTTGCTGCTTTTAAAGAGACAAATTTATGAGAGCCAAAGACATTTGAAATTTAT	300
Dp	967	GGAAGATATACGCCGTTTTTAAAGAGACAGCTGTTTAAAGCCAAAGAACATTTGAAATTTAC	1026
Qy	301	ANGAAGCAAACTAAAAATGCTCCAGGAAGACATCATGAATACTTAAGAGAGAGATTGAA	360
Dp	1027	ANGAAGACAGACAGAAATGCTCTGGGGAGAGATCATGAATCTTGAGAGAGAGATTGAA	1086
Qy	361	AATGAGACTAAAGAGCTTGTTTTTTTCTACAAAGTAGATTGAGAAATTTAAAGATTTA	420
Dp	1087	AATGAGACTAAAGAGCTCTGTTTTTTCTTACAGAGAGTAATTGAAGAAATTTAAAGACTTA	1146
Qy	421	GAAGAAATGACTCCCAAGACATGCAAGATGTAATTTCTATCAGATTTGGGACATCATGAA	480
Dp	1147	GAAGAAATGAACTCCCAAGACATGCAAGATGTAATTTCTTGTGATTTAGGACATCATGAA	1206
Qy	481	AGGCTCTTAATGAGCGATCTATCTACCTCAGCAAAACAGATGGGGGACAGGATGGCGT	540
Dp	1207	AGGCTCTTAATGAGCGATCTATCTACTCTCAGTACAGACAGATGAGCAGGATGGCGG	1266
Qy	541	GAAGAAGAGCCCAAAAGATCTGACAGAGCTGGTCCAGCCGAGAAATTAATCTTCAGAT	600
Dp	1267	GAAGAAGAGGCCCAAAAGATCTGACAGAACTGGTTACAGCCGAGAAATTAATCTTCAGAT	1326
Qy	601	CCCAAGACATGACAGCAAAAGCAAGCTATGTTGTAATATCACAAAGCTGTGGCTAT	660
Dp	1327	CCCAAGACATGACAGCAAAAGCCAAAGAGCTGTGTAATATCACAAAGGCTGTGGCTAT	1386
Qy	661	GGCGTACAGCCCATCTGATGATGTCGTCGCTTATGATTTGGCATTTGGACCCAGGAA	720
Dp	1387	GGCGTACAGCCCATCTGATGATGTCGTCGCTTATGATTTGGCATTTGGACCCAGGAA	1446
Qy	721	CTCGCCTGGCAATCTCAATTTGGCCCTACGCGTACTGGGGGATGGGAAACTGTGTTTGA	780
Dp	1447	CTCATCTTTGGATCTCGAATTTGGCCCTATGCTAGTGGATGGGAGACGTGTAATTTAG	1506
Qy	781	CCTGTAACTGAGACGTGCACAGACAGATCTGGCAGCTCCACTGTGACATTTGGTCAGGTAA	840
Dp	1507	CCTGTAACTGAGACATGCACAGACAGATCTGGCATCTCCACTGTGACACTGGTCAAGGTAA	1566
Qy	841	GTTAAAGCAAAAATCTTTCAGGTGGTGGAGCTCCCATTTGTAGACAGTTCATCTGCT	900
Dp	1567	GTGAAGACAAAATCTTTCAGAGTGGTGCAGCTTCCCATTTGTAGACAGTTCATCTCCGT	1626
Qy	901	CCCTCATATTTTACCCTCGCTGTCCAGAGAAGACTTTGCAGATCGACTTGTACAGTCCAT	960
Dp	1627	CCCTCATATTTTACCCTTGGCTGTACCAAGAAAGCCTGCAGATCGACTTGTACAGTCCAT	1686
Qy	961	GCTATCTCTGCAGTGTGGTGGTATCCCACTTTGTCAAGTACTTGAATTTGCCACACACC	1020
Dp	1687	GCTATCTCTGCAGTGTGGTGGTCTCTCAAGTTGTCAAACTCTTATCCGCCACAGCCT	1746
Qy	1021	TGGCTGGAAAAAGAAATAGAAAGAGGCCACCAAGAACTGTGGCTCAAACTCCAGTTAT	1080
Dp	1747	TGGCTAGAAAAAGAAATAGAAAGAGGCCACCAAGAACTGTGGCTCAAACTCCAGTTAT	1806
Qy	1081	GGAGCTCATGTTTAGACAGCAGACAGAAAGTGGAGCGGAAGACAGCTTCATCCATTGAG	1140
Dp	1807	GGAGCTCATGTCAGAGCCACAGACAAAGTGGAGAACGAAGCTGCTTCATCCATTGAA	1866
Qy	1141	GAATACAGGTGACAGCTTTGAAGAGACTTTTCAGCTTTTGCTGCAGAAATGCAAGTGGAT	1200
Dp	1867	GAGTACATGTGGCTATTTGAAGAACTTTTTCAGCTTCTTGACGCCAGAAATGCAAGTGGAC	1926
Qy	1201	AAAAAAGCGGTATTTTGGCCACAGATGACCCCTCTTTATTAAGAGAGCCAAAAAAG	1260
Dp	1927	AAAAAAGAGGTATTTTGGCCACAGATGACCCCTCTTTATTAAGAGAGCCAAAAAAG	1986
Qy	1261	TACCCCAAGTTAGAAATTTATAGTAGTAATCTATCTCTTGGTCAAGCTGCAGATACATAT	1320
Dp	1987	TACCCCAATTTATGAATTTATTAAGTAGTAATCTATTTCTTGCTCAAGCTGCAGATACAT	2046
Qy	1321	CGATATACAGAAATTCATCTTGGGGGTGATCTTGATATACACTTCTCTCCAGCA	1380

Db	2047	CGATACACAGAAATTCCTCTGCGATGATCCTGGATATACATTTCTCTCAGGCA	2106
Qy	1381	GACTTCTAGTGTGTAATTTTTCATGCGAGGCTGTGAGATTGCTTATGAATCATGCA	1440
Db	2107	GACTTCTAGTGTGTAATTTTTCATGCGAGGCTGTGAGATTGCTTATGAATCATGCA	2166
Qy	1441	GGCGTCATCCTGTATGCTCTGCGAAGCTTCGTTCTTTTGATGACATCTATTTTGG	1500
Db	2167	ACATCTACATCCTGTATGCTCTGCGAAGCTTCGTTCTTTTGATGACATCTATTTTGG	2226
Qy	1501	GGCCCAATGGCCCAACCAATTTGGCATTTTATTCCTCACCAACCTCGAAGTGAAGGAA	1560
Db	2227	GGCCCAATGGCCCAACCAATTTGGCATTTTATTCCTCACCAACCTCGAAGTGAAGGAA	2286
Qy	1561	ATCCCCATGGAACCTGGAGATATTATTGTGTGTCGTCGTAATCTAGTGGATCTTCT	1620
Db	2287	ATTCCTCATGGAACCTGGAGATATTATTGTGTGTCGTCGTAATCTAGTGGATCTTCT	2346
Qy	1621	AAAGGTGTAAACAGAAAACTGGAGAGAGGCTATATCTCTCTACAAAGTTGAGAG	1680
Db	2347	AAAGGTGTAAACAGAAAACTGGAGAGAGGCTATATCTCTCTACAAAGTTGAGAG	2406
Qy	1681	AAGTATGAACAGTCAGTATCCCAATATCCCGAGGCTGACAACTAA	1728
Db	2407	AAGTATGAACAGTCAGTATCCCAATATCCCGAGGCTGACAACTAA	2454
RESULT 5			
ABV28283			
Id	ABV28283	standard; cDNA; 3291 BP.	
xx	ABV28283;		
xx	16-SEP-2002	(first entry)	
xx	Human prostate expression marker	cDNA 28274.	
xx	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	pharmacogenomic marker; gene; ss.	
xx	Homo sapiens.		
xx	MO200160860-A2.		
xx	23-AUG-2001.		
xx	20-FEB-2001;	2001MO-US05171.	
xx	17-FEB-2000;	2000US-183319P.	
xx	16-MAR-2000;	2000US-189862P.	
xx	25-MAY-2000;	2000US-207454P.	
xx	09-JUN-2000;	2000US-211314P.	
xx	18-JUL-2000;	2000US-219007P.	
xx	13-DEC-2000;	2000US-255281P.	
xx	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
xx	Schlegel R, Endege WO, Monahan JE;		
xx	WPI: 2001-662795/76.		
xx	Novel isolated nucleic acid molecule associated with cancerous state of		
xx	prostate cells and correlating with presence of prostate cancer, useful		
xx	for detecting presence of prostate cancer, stage of prostate cancer -		
xx	Claim 1; Page 5883; 11750pp; English.		
xx	The invention relates to an isolated nucleic acid molecule (I) comprising		
xx	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the		
xx	specification or its complement. (I) is useful for:		
xx	(a) assessing whether a patient is afflicted with prostate cancer;		
xx	(b) monitoring the progression of prostate cancer in a patient;		
xx			





xx	Homo sapiens.
os	WO200122920-A2.
xx	05-APR.-2001.
pn	28-SEP-2000; 2000MO-US26524.
xx	29-SEP-1999; 99US-0157137.
pd	03-NOV-1999; 99US-0163280.
xx	(HUMA-) HUMAN GENOME SCI INC.
pf	Ruben SM, Barash SC, Birse CE, Rosen CA;
xx	WPI: 2001-235357/24.
pi	P-PsDB: AAG73884.
dr	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
xx	useful for preventing, diagnosing and/or treating colorectal cancers -
pt	Claim 1; Page 2480-2481; 9803pp; English.
xx	AAB32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
xx	cancer-associated nucleic acid molecules (N) and proteins (P), where
xx	the proteins are collectively known as colon cancer antigens. The colon
xx	cancer antigens have cytostatic activity and can be used in gene
xx	therapy and vaccine production. N and P may be used in the prevention,
xx	diagnosis and treatment of diseases associated with inappropriate P
xx	expression. For example, N and P may be used to treat disorders
xx	associated with decreased expression by rectifying mutations or deletions
xx	in a patient's genome that affect the activity of P by expressing of P.
xx	Inactive proteins or to supplement the patients own production of P.
xx	Additionally, N may be used to produce the colon cancer-associated Ps,
xx	by inserting the nucleic acids into a host cell and culturing the cell
xx	to express the proteins. N and P can be used in the prevention, diagnosis
xx	and treatment of colorectal carcinomas and cancers. AAB37196 to AAH37204
xx	and AAB77789 represent sequences used in the exemplification of the
xx	present invention.
cc	N.B. Pages 666 to 682 and page 7053 of the sequence listing were
cc	missing at time of publication, meaning no sequences are present for
cc	SEQ ID NO:1027 to 1052, 7921 and 7922.
xx	Sequence 3007 BP; 900 A; 597 C; 668 G; 834 T; 8 other;
sq	
Query Match	86.7%; Score 1499; DB 22; Length 3007;
Best Local Similarity	92.0%; Pred. No. 0;
Matches 1591; Conservative	1; Mismatches 136; Indels 1; Gaps 1
Oy	1 ATGCGGCATGGACGTCTTGTCGTCGTTGGATTATGCTCATTTTGGCTGGGACC 60
Dd	
	420 ATGCGGCATGGACGTCTTGTCGTCGTTGGATTATGCTCATTTTGGCTGGGACC 479
Oy	61 TTGCTATTTTACATAGTGTGTCACTGTGTGAAGATAAATACACACTGTGACTAGC 120
Dd	
	480 TTGCTATTTTATATAGTGTGTCACTGTGTGAAGATAAATACACACTGTGACTAGC 539
Oy	121 CGAGAAGTCTCCAAAGATTTTGGCAAAGCTGGAACGCTTAAMACAACAAATGAAGATTG 180
Dd	
	540 CGAGAAGTCTCCAAAGATTTTGGCAAAGCTGGAACGCTTAAMACAAGCATGAAGATTG 599
Oy	181 AGGAGATGCGCTG-GATCTCTCCGAATPACCAAGAGGCCCATTTGATCAGGGCCAGCTTC 239
Dd	
	600 AGGAGATGCGCGAAATCTCTCCGATACAGAAAGGCCCATTTGATCAGGGCCAGCTAT 659
Oy	240 AGGAGATGCTGCGCTTTGAAGAGCAATTTATGAAGGCCCAAGAACAGATTGAAAATTGA 299
Dd	
	660 AGGAGATGACGGCTTTTGAAGAGCAGACTTTTAAAGGCCCAAGAACAGATTGAAAATTGA 719
Oy	300 TAAGAAACAACCTAAAAATGCTCAGAGGAAGATCATGAATCTTAAGAGAGGAGATTGA 359
Dd	
	720 CAAGAAACAGACAGAAATGCTGTGGGAAGATCATGAATCTGTGAGAGAGGAGATTGA 779
Dd	

Qy	360	AAATGAGACTAAAGAGCTCTGTTTTTCTTACAAAGTGAAGACTTGAAGAATTTAAACATATT	419
Db	780	AAATGAGACTAAAGAGCTCTGTTTTTCTTACAAAGTGAAGACTTGAAGAATTTAAACATATT	839
Qy	420	ACAAGGAATGAACCTCCAAAGACATGCAGATGGAATTTCTATCAGATTTGGACATCATGA	479
Db	840	ACAAGGAATGAACCTCCAAAGACATGCAGATGGAATTTCTTGTGATTTAGGACATCATGA	899
Qy	480	AAGGTCATTAATGACGGATCTATTACTACCTACGTCAAACAGATGGGGCAGTGATTTGGC	539
Db	900	AAGGTCATTAATGACGGATCTATTACTACCTACGTCAGACAGATGGAGCAGTGATTTGGC	959
Qy	540	TGAAGAGGAGCCAAAGATCTGCAGAGCGTGGCCAGCGAGATTAACATATCTTCAGAA	599
Db	960	GGAAGAAAGAGCCAAAGATCTGCAGAGACGTGGTTCAGCGAGATTAACATATCTTCAGAA	1019
Qy	600	TCCCAAGAGCTGCAGCAAAAGCCAAAGAGCTAGTGTTAATATCAACAAAGCCTGTGGCTA	659
Db	1020	TCCCAAGAGCTGCAGCAAAAGCCAAAGAGCTAGTGTTAATATCAACAAAGCCTGTGGCTA	1079
Qy	660	TGGCTGTACGCTCCATCATGTAGTGTACTGCTTTATGATTTGCATATGGCACCCAGCCAAAC	719
Db	1080	TGGCTGTACGCTCCATCATGTAGTGTACTGCTTTATGATTTGCATATGGCACCCAGCCAAAC	1139
Qy	720	ACTGCGCTTGGAAATCTCAATTTTGGCCCTACCGTACGAGGGGGTGGGAAACGTGTATTAG	779
Db	1140	ACTCATCTTTGGAAATCTCAAGATTTGGCCCTATGCTACTGCGATGGGAGACTGTATTATAG	1199
Qy	780	ACCTGTAAAGTAGAGACGTGCACAGACAGATCTGGCAGCTCCACTGACATTTGGTCAGGTGA	839
Db	1200	GCGTTGTAGTAGAGACGTGCACAGACAGATCTGGCATCTCCACTGACATCTGGTCAGGTGA	1259
Qy	840	AGTAAAGGACAAAATGTCTCAGGTGGTGGAGTCCCATTTGTAGACAGTGTTCATCTCG	899
Db	1260	AGTGAAGGACAAAATGTCTCAGGTGGTGGAGTCCCATTTGTAGACAGTGTTCATCTCGCCG	1319
Qy	900	TCCCTCATATTATACCCCTGGCTGTGCCAGAAAGCCTTGCAGATCGACTGTACGAGTCCA	959
Db	1320	TCCCTCATATTATACCCCTGGCTGTACCAGAAAGCCTGCAGATCGACTGTACGAGTCCA	1379
Qy	960	TGGTATATCTCTCAGCTGTGTGGGTATTCGCCAGTTTGTCAAGTACTTGATTTCCGCCCAACAC	1019
Db	1380	TGGTATATCTCTCAGCTGTGTGGGTATTCAGTTTGTCAAAATCTGTGATTCGCCCAACAC	1439
Qy	1020	CTGCGCTGGAAGAGGAATTAGAAGAGGCCACCAAGAGCTTCAAAATCATCCAGTTAT	1079
Db	1440	TTTGCTTGAAGAAAGAAATTAGAAGAGGCCACCAAGAGCTTGGCTTCAAAATCATCCAGTTAT	1459
Qy	1080	TGGAGTCCATGTTAGAGCGACAGACAAAGTGGGAGCGGAGACGCTTCCATGCCATTTGA	1139
Db	1500	TGGAGTCCATGTTAGAGCGACAGACAAAGTGGGAGACGAGTGGCTTCCATGCCATTTGA	1159
Qy	1140	GGAATACACGGGTGCAGCTGTGAAGAAAGCTTTCAGCTTTTGGCTGGCAGAAATCCAGTGTGA	1199
Db	1560	AGAGTACATGGTGCATGTGGAAGAACATTTTCAGCTCTTTGACACGCAAGATTCAGTGTGA	1619
Qy	1200	TAAAAAAGAGGTGTATTGGCCACAGATGACCCCTGCTTTGTTAAAAAGAGCAAAAAACAA	1259
Db	1620	CAAAAAAGAGTGTATTGGCCACAGATGACCCCTTATTAAAAAGAGCAAAAAACAA	1679
Qy	1260	GTACCCCAATTATGAATTTATTAGTGAATCACTTAATTTCTGGTCAAGCTGCATCCACAA	1739
Db	1680	GTACCCCAATTATGAATTTATTAGTGAATCACTTAATTTCTGGTCAAGCTGCATCCACAA	1759
Qy	1320	TGCAATATCAGAAATTCACGTGGGGGTGTGATCTCGAATATACATTTCTCTCCAGGC	1379
Db	1740	TGCAATACAGAAATTCACCTCTGTGAGTGTGATCTCGAATATACATTTTCTCTCTCAGGC	1799
Qy	1380	AGACTTCCTAGTGTATCTTTTTCATGCGAGGTCTGTAGAGTTGCTTATGAATATCATCA	1439
Db	1800	AGACTTCCTAGTGTATCTTTTTCATGCGAGGTCTGTGACAGTTGCTTATGAATATCATCA	1859







```

OY 541 GAAAGAGGCCAAGATCTGACAGAGCTGTCCAGCGGAGATATCATCTTCAGAAAT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 541 GAAAAAGAGGCCAAGATCTGACAGAGCTGTCCAGCGGAGATATCATCTTCAGAAAT 600
OY 601 CCCAAGAGCTGACGAAAGCCAAAGAGCTAGTGTGTAATATCAACAAGGCTGTGCTAT 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 601 CCTAAGAGCTGACGAAAGCCAAAGAGCTGTGTGTATCATCAATTAAGGCTGTGCTAT 660
OY 661 GGCTGCTACCTCCATCATATGATGTACTGTCTTATATGATTTGCAATGCGCCAGCAACA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 GGTTGTCAACTCCATCTGATGTCTGTACTGTTCATGATTTGTTATGTCACCCAGCCAAACA 720
OY 721 CTGCGCTTGGAAATCTCACATTTGGCGCTACGCTACTGGGGGATGGGAAACTGTGTTAGA 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 CTGATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGTGTGATGGAGACTGTGTTAGA 780
OY 781 CCTGTAAGTAGAGACGTGACAGACAGATCTGGCAGCTCCACTGACATTTGGTCAGGTGAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 CCTGTAAGTAGAGACGTGACAGACAGATCTGGCGCTCTCCACTGACATGTGTGAGGTGAA 840
OY 841 GTAAAGGACAAAAAATGTTACAGGTGTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 GTAAATGACAAAAACATTTCAAGTGTGTGAGCTCCCATTTGTAGACAGCTTCCTCGG 900
OY 901 CCTCCATATTTACCCCTGGCTGTCCAGAAAGACCTTGACAGATGACATTTGACGAGTCCAT 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 901 CCTCCTTACTTACCACTGGCTGTGTCAGAAAGACCTTGACAGCGACTCCTTAAGAGTCCAT 960
OY 961 GGTGATCTGACAGTGTGTGGGTATCCAGTTTGTCAAGTACTGTGATTCGCCCAACCC 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 961 GGTGACCTCTGACAGTGTGTGGGTGTCCAGTTTGTCAAAATGTTGATGTGTCCACAAACGT 1020
OY 1021 TGGCTGGAAAAAGAAATAGAAAGGCCACCAAGAACTAGCTTCAACATCCAGTTAT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1021 TGGCTGGAAAAAGAAATAGAAAGGCCACCAAGAACTAGCTTCAACATCCAGTTAT 1080
OY 1081 GGAGTCCATGTTAGAGCAGACAGAAAGTGGGAGGAGCAGCCCTCCATCCCATTTGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1081 GGAGTCCATGTTAGAGCAGACAGAAAGTGGGAGGAGCAGCCCTCCATCCCATTTGAG 1140
OY 1141 GAATACAGCGTGCACGTTGAGAAAGACTTTCAGCTTCTTGCTGCGAAATGCAAGTGAT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1141 GAGTACATGTTACACGTTGAGAAACTTTCAGCTTCTTGCTGCGAAATGCAAGTGAT 1200
OY 1201 AAAAAAAGGTGTATTTGGCCACAGATGACCCCTGTTGTTAAAGAGGCAAAAAACAAG 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1201 AAAAAAAGGTGTATTTGGCCACAGATGACCCCTGTTGTTAAAGAGGCAAAAAACAAG 1260
OY 1261 TACCCAGTTATGAATTTTATAGTATACTCTATCTTGGTCAGCTGACATACATTAAT 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1261 TACCCAGTTATGAATTTTATAGTATACTCTATCTTGGTCAGCTGACATACATTAAT 1320
OY 1321 CGATATACAGAAAAATTCACCTGGGGTGTGATCCTGATATACACTTTCCTCCAGGCA 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1321 CGATATACAGAAAAATTCACCTGGGGTGTGATCCTGATATACACTTTCCTCCAGGCT 1380
OY 1381 GACTTCTGAGTGTGATCTTTTATGCGAGGTGTGATGAGTTGCTTATGAATCATGCAA 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1381 GACTTCTGAGTGTGATCTTTTATGCGAGGTGTGATGAGTTGCTTATGAATCATGCAA 1440
OY 1441 GGCGTGCATCTGATGCTCTGCGAACTTCGCTTGGATGACATCTACTATTTTGA 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1441 ACCCTCATCTGATGCTCTGCGAACTTCGCTTGGATGACATCTACTATTTTGA 1500
OY 1501 GGCCAAATATGCCCAATCAGATGCTGCTTATCTCACAAACCTGCAACTGAAGAGAA 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1501 GGCCAAATATGCCCAATCAGATGCTGCTTATCTCACAAACCTGCAACTGAAGAGAA 1560
OY 1561 ATCCCATGGAACCTGGAGATATATTGTTGGTGTGCGAAATACATGCGATGCTATCC 1620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1561 ATTCATGGAACCTGGAGATATATTGTTGGTGTGCGAAATACATGCGATGCTATCC 1620
OY 1621 AAGGTGTTAACAGAAAACTGGGAAGAGCGGCTATATCCCTCTTAACAAAGTTGAGAG 1680

```

```

DB 1621 AAGGTGTTAACAGAAAACTGGGAAGAGCGGCTATATCCCTCTTAACAAAGTTGAGAG 1680
OY 1681 AAGATGGAAGAGTCAAGTACCCACATATCCCGAGGCTGACAGTA 1727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1681 AAGATGGAAGAGTCAAGTACCCACATATCCCGAGGCTGACAGTA 1727
RESULT 9
AA87952
ID AA87952 standard; DNA; 1017 BP.
XX
AC AA87952;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.
XX
KW Human: alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunosay;
KW ds.
XX
OS Homo sapiens.
XX
FH Key 1.1017 location/Qualifiers
FT CDS /partial a
FT /product= "alpha 1-6 fucosyltransferase 237-575"
FT JP2001011097-A.
XX
PD 16-JAN-2001.
XX
PF 29-JUN-1999: 99JP-0183569.
XX
PR 29-JUN-1999: 99JP-0183569.
XX
PA (FJRE ) FUJIREBIO KK.
XX
DR WPI: 2001-275926/29.
XX P-PSDB: AAB75061.
XX
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunosay.
XX
PS Example 1; Page 6-7; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunosay for detecting human
CC alpha 1-6 fucT by using the above antibody of its antibody fragment;
CC and (3) a reagent used for the above immunosay. The anti-human
CC alpha 1-6 fucT antibody can be used for immunosay. The present
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to
CC 575, which is used in an example from the present invention.
XX
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;
Query Match 50.8%; Score 878.4; DB 22; Length 1017;
Best Local Similarity 91.5%; Pred. No. 5.6e-241;
Matches 930; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
OY 709 ACCCAGCAACTCGCTTGAATCTCACAAATTTGGCGTACCTACTGGGGATGGGAA 768
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 ACCCAGCAACTCGCTTGAATCTCACAAATTTGGCGTACCTACTGGGGATGGGAG 60
OY 769 ACTGTGTTAGACCTTAAAGTACGACGACAGACATCTGGCAGCTCCACTGCACAT 828
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 ACTGTATTAAGGCTCTAAGTACGACGACAGACATCTGGCAGCTCCACTTGGACAC 120
OY 829 TGGTCAGGTGAAGTAAAGCAAAAAATGTTCAAGTGTGAGCTCCCATTTGAGACAGT 888

```

```
|||||
Db 121 TGGTCAGGTGAAGTGAAGACAAATGTTCAGAGTGTGAGCTTCCATTGTAGACAGT 180
Oy 889 GTTCATCTGTCCTCCATATTTACCCCTGCTGCTCCAGAAAGACTTGCAGATTCACATT 948
Db 181 CTTTCATCCCTGCTCTCATATTTTACCTTGGCTGTACAGAAACCTTCGAGATGCACATT 240
Oy 949 GTACGAGTCATGTCATCTGTCAGTGTGGGTATCCAGTTTCTCAAGTACTTGATT 1008
Db 241 GTACGAGTCATGTCATCTGTCAGTGTGGGTGTCTGCTCAGTTTGTCAATATCTTGATC 300
Oy 1009 CGCCACAAACCTGCTGGAAGAAAGAAATGAAAGAGGCCACCAAGAGCTTAGCTTCAA 1068
Db 301 CGCCACACACCTTGGCTAGAAAAAATGAAAGAGCCACCAAGAGCTTGGCTTCAA 360
Oy 1069 CATCCAGTTATTTGAGATGTCATGTTAGACGACAGCAAAAGTGGAGCGGAGAGGCTTC 1128
Db 361 CATCCAGTTATTTGAGATGTCATGTTAGACGACAGCAAAAGTGGAAACAGAGGCTTC 420
Oy 1129 CATCCAGTTGAGAAATACACGCTGACGCTTGAAGAGACTTTCAGCTTCTGCTCCAGA 1188
Db 421 CATCCAGTTGAGAAATGATGTCATGTTGAAAGAACTTTTACGCTTCTGTCAGCAGAG 480
Oy 1189 ATGCAAGTGGATTAAGAAAGGCTGATTTGGCCACAGATGACCTTCTTGTAAAGAG 1248
Db 481 ATGCAAGTGGACAAAAAAGAGTGTATTTGGCCACAGATGACCTTCTTGTAAAGAG 540
Oy 1249 GCAAAAACAAGTACCCAGTTGATGAAATTTATGATGATACCTTCATCTGTCAGCT 1308
Db 541 GCAAAAACAAGTACCCAGTTGATGAAATTTATGATGATACCTTCATCTGTCAGCT 600
Oy 1309 GCATCATATATGATATACAGAAATTCATCTCGGGGTGTGATCCTGATATACACTTT 1368
Db 601 GGATGTCACATGATGATACAGAAATTCATCTCGTGGAGTATCCTGGATATACACTTT 660
Oy 1369 CTCTCCAGGACACTTCTGATGTTCTATCTTTTCAATCCAGGCTGTGATAGCTTAT 1428
Db 661 CTCTCTCAGGACACTTCTGATGTTCTATCTTTTCAATCCAGGCTGTGATAGCTTAT 720
Oy 1429 GAAATGATGAAAGCTGATCTGATGCTGCGAAGCTTCCGTTCTTGGATGACATC 1488
Db 721 GAAATGATGAAACACTACATCTGATGCTGCGAAGCTTCCATCTTGTAGATGACATC 780
Oy 1489 TACTATTTTGAAGCCCAATGCCCACAAACCAATTTGCCATTTCTGACCAACCTGCA 1548
Db 781 TACTATTTTGGGGCCAGATGCCCACAAATTCATTAATGCTATCACAACCCGA 840
Oy 1549 ACTGAAGAGAAATCCCATGAGACCTGGAATATTTATGCTGTGCTGGAATTCACCTG 1608
Db 841 ACTGAGATGAATTTCCCATGGAACCTGGAGATTCATTTGCTGTGGAATTCATTTG 900
Oy 1609 GATGGCTATCTTAAAGTTTACAGAAAACCTGGGAAGGCGGCTATATCCCTCTAC 1668
Db 901 GATGGCTATCTTAAAGTTTACAGAAAATTTGGGAAGGCGGCTATATCCCTCTAC 960
Oy 1669 AAGTTTCGAGAGAGATAGAAACAGTCAAGTACCCCAATATCCGAGGCTGACAA 1724
Db 961 AAGTTTCGAGAGAGATAGAAACAGTCAAGTACCCCAATATCTTGAAGGCTGAGAA 1016

RESULT 10
AAC63891
ID AAC63891 standard; cDNA; 979 BP.
XX
AC AAC63891:
XX
DT 09-FEB-2001 (first entry)
XX
DE Chinese hamster FUT8 cDNA.
XX
KW Immunologically functional molecule; immune system; immunomodulation;
glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;
viral infection; bacterial infection; allergy; autoimmune disease;
KW
```

```
KW Inflammation; antibody; Chinese hamster; FUT8; ss.
XX
OS Cricetus griseus.
PN WO200061739-A1.
PD 19-OCT-2000.
PE 07-APR-2000; 2000MO-JP02260.
PR 09-APR-1999; 99JP-0103158.
XX (KYOW ) KYOMA HAKKO KOYO KK.
PI Hanei N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;
DR WPI: 2000-665129/64.
XX
PT Control of activity of antibodies and other immunologically functional
PT molecules by addition or removal of fucose from sugar chain for
PT diagnosis and treatment of cancer, allergy and other diseases
XX
PS Example 8; Page 72-73; 81pp; Japanese.
XX
CC The invention relates to a method for controlling the activity of an
CC immunologically functional molecule (e.g., an antibody) where the
CC control is effected by the presence or absence of fucose bound to an
CC N-acetylglucosamine residue at the reducing end of the sugar chain on
CC the immunologically functional molecule. The invention also relates to
CC methods for the diagnosis, prevention or treatment of diseases which
CC involve the modified immunologically functional molecule, and agents
CC which stimulate the activity of an immunologically functional molecule.
CC The methods of the invention are used for the diagnosis, treatment and
CC prevention of a broad range of diseases including cancer, circulatory
CC disease, viral or bacterial infection, allergy, autoimmune disease and
CC inflammation. The present sequence represents a Chinese hamster FUT8
CC cDNA isolated in an exemplification of the invention.
XX
SO Sequence 979 BP; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 48.2%; Score 833.4; DB 21; Length 979;
Best Local Similarity 90.7%; Pred. No. 4,3e-228;
Matches 888; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Oy 720 ACTGCGCTTGATGATTCACAAATTTGGGCTACGCTACCTGGGGGATGGGAACGTTGAG 779
Db 1 ACTCATCTTGGATCTCAGATTTGGCGTATGCTACTGGAGATGGAGACTGTGTTAG 60
Oy 780 ACCTGTAAGTGAAGCTGACACAGACAGATCTGACACCTCCACTGACATTTGGTCAAGTGA 839
Db 61 ACTGTAAGTGAACATGACACAGACAGAGTCTGGCTCTCCACTGAGACACTGTCAAGTGA 120
Oy 840 ACTGAAGCAAAATGTTCAAGTGTGAGTCTGCCATTTGTAGACAGTGTATCTCTG 899
Db 121 AGTGAAGCAAAATGTTCAAGTGTGAGTCTGCCATTTGTAGACAGCTTCATCTCTG 180
Oy 900 TCTCATATTTTACCCCTGCTGCTCCAGAAAGCTTGGAGATGCACTTGTACAGATGCA 959
Db 181 TCTCTCTTACTTACCTTGGCTGTACAGAGACCTTGTGAGATGCACTTGTGAGATGCA 240
Oy 960 TGGTGATCTGCAAGTGTGCTGATCCAGTTTGTCAAGTACTTGTATCCGCCACAC 1019
Db 241 TGGTGATCTGCAAGTGTGCTGATCCAGTTTGTCAAGTACTTGTATCCGCCACAC 300
Oy 1020 CTGCTGGAAGAAATGAAAGAGAGCCACCAAGAGCTTAGCTTAAACATTCAGTTAT 1079
Db 301 TTGGCTGGAAGGAAATGAAAGAAACCAAGAGCTTGGCTTCAACATTCAGTTAT 360
Oy 1080 TGGAGTCCATGTTAGACGACAGCAAAAGTGGAGGGAAGCAAGCTTCATCCCACTTGA 1139
Db 361 TGGAGTCCATGTTAGACGACAGCAAAAGTGGAGGGAAGCAAGCTTCATCCCACTTGA 420
```

Oy	1140	GGATATACACGGTGCACGCTTGAAAGAACTTTCAGCTTCTTGCTCGAGAAATGCAGTGGAA	1139
Db	421	GGAAATACATGGTGTACACGCTTGAAAGAACTTTCAGCTTCTCGAAACGGCAAGATGAAAGTGA	480
Oy	1200	TAAAAAAAGGCTGATTTGGCGACAGATGACCCCTGCTTGTGTTAAAGAGGCAAAAAACAA	1255
Db	481	TAAAAAAGAGACTGATCTGGCGACAGTATGACCCTTCTTTGTTAAAGGAGGCAAAAGACAAA	540
Oy	1260	GTACCCCAAGTTATGAATTTATTAAGTAACTATCTCTTGCTGACGTGCACATATA	1315
Db	541	GTACCCCAATTTATGAATTTATTAAGTAACTATCTCTTGCTGACGTGCACATACAAA	600
Oy	1320	TGCAATATACAGAAATTCACCTGGGGGTGATCCTGGATATACACTTCTCCAGGC	1375
Db	601	CCGATACACAAAAATTCACCTGGGGCGGATCCGGATATACACTTCTCTCCAGGC	660
Oy	1380	AGACTTCCTAGTGTACTTTTCATGGCAGGCTGTAGAGTGCCTTATGAATATCATGA	1435
Db	661	TGACTTCTTGTGTACTTTTCATGCCAGGCTGTAGGGTTGCTTATGAATCATGCA	720
Oy	1440	AGCGTGCATCTCATATGCTCTGCGAATTCGCTTCTTTGGATGACATCTACTATTTTGG	1495
Db	721	AACACTGCATCTGATGCTCTGCAAACTTCATCTTTAGATGACATCTACTATTTTGG	780
Oy	1500	AGGCCCAATATCCCAACAACCAATTGCCATTTTCTCTCACCACTCGACTGAAAGAGAA	1555
Db	781	AGGCCCAAAATCCCAACAACCGATTTGCGATTTTCTCTCACCAACTCGAATTAAGAGGAA	840
Oy	1560	AATCCCATGGAACCTGAGATTAATTATGGTGGGCGGAAATCACTGGGATGGGCTATCC	1615
Db	841	AATCCCATGGAACCTGAGATTAATTTATGGTGGGCGGAAACCAATTTGAATGATGTTACTC	900
Oy	1620	TAAAGGTGTTAACAGAAAACTGGGAGAGACGGGCTATATCCCTCTACAGAAATTTGAGA	1675
Db	901	TAAAGGTGTCAACAGAAAACTAGSAAAAACAGGCGCTGTACCTCTCTACAAAGTCGAGAA	960
Oy	1680	GAGATAGAAACACTCAAG	1698
Db	961	GAGATAGAAACGCTCAAG	979

XX	RESULT 11
XX	ABK70036
XX	ID ABK70036 standard; CDNA: 979 BP.
XX	ABK70036;
XX	15-JUL-2002 (first entry)
XX	Antibody production method related CDNA #4.
XX	Antibody production: cytostatic; immunomodulator; vasotropic; virucide;
KW	antibacterial; antiinflammatory; antiallergic; allergy; inflammation;
KW	autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO
XX	tumour; circulatory disease; infection; primer; ss.
OS	Cricetulus griseus.
XX	
PN	WO200231140-A1.
XX	
PD	18-APR-2002.
XX	
PF	05-OCT-2001; 2001WO-JP08804.
XX	
PR	06-OCT-2000; 2000JP-0308526.
XX	
PA	(KYOM ) KYOMA HAKKO KOGYO KK.
XX	
PI	Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
XX	Hosaka E, Yamano K, Yamasaki M, Hanai N;
DR	WPI: 2002-340182/37.
XX	

[illegible]

Db 661 TGACTCTCTGTGTGACCTTTTTCATCCAGGTCGTGTAGGTTGCTTANGAATCATGCA 720  
Qy 1440 AGCGTCGATCTGATAGCTCTGCAACTCCGTTCTTTGATGACATCTACTATTTGG 1499  
Db 721 AACACTGCACTCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTGG 780  
Qy 1500 AGGCCCAAAATGCCCCACACCAAAATTCGCAATTCCTCACCACCTCGAAGTGAAGAGA 1559  
Db 781 AGGCCCAAAATGCCCCACACCAAAATTCGCAATTCCTCACCACCTCGAAGTGAAGAGA 840  
Qy 1560 AATCCCCATGGAACCTGAGATATTTATGTGTGCTGGAAATCACTGGATGGCTATCC 1619  
Db 841 AATCCCCATGGAACCTGAGATATTCATTTGTGTGCTGGAAATCACTGGATGGCTATCC 900  
Qy 1620 TAAAGGTGTTAAAGAAACTGGAGAGAGGCGCTATATCCCTGACAAAGTTGAGAGA 1679  
Db 901 TAAAGGTGTTAAAGAAACTGGAGAGAGGCGCTATATCCCTGACAAAGTTGAGAGA 960  
Qy 1680 GAAGATAGAAACAGTCAG 1698  
Db 961 GAAGATAGAAACAGTCAG 979  
  
RESULT 12  
AAC63892  
ID AAC63892 standard; cDNA: 979 Bp.  
XX AAC63892:  
AC AAC63892:  
DT 09-FEB-2001 (first entry)  
XX  
DE Rat FUR8 cDNA.  
XX  
KM Immunologically functional molecule; immune system; immunomodulation;  
KM glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;  
KM viral infection; bacterial infection; allergy; autoimmune disease;  
KM inflammation; antibody; rat; FUR8; ss.  
XX  
OS Rattus sp.  
XX  
PN MO200061739-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 07-APR-2000; 2000WO-JP02260.  
XX  
PR 09-APR-1999; 99JP-0103158.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;  
PI Imabepu S, Kanda Y, Yamane N, Anazawa H;  
XX  
DR WPI; 2000-665129/64.  
XX  
XX  
PT Control of activity of antibodies and other immunologically functional  
PT molecules by addition or removal of fucose from sugar chain for  
PT diagnosis and treatment of cancer, allergy and other diseases  
XX  
PS Example 8; Page 73-74; 81pp; Japanese.  
XX  
CC The invention relates to a method for controlling the activity of an  
CC immunologically functional molecule (e.g., an antibody) where the  
CC control is effected by the presence or absence of fucose bound to an  
CC N-acetylglucosamine residue at the reducing end of the sugar chain on  
CC the immunologically functional molecule. The invention also relates to  
CC methods for the diagnosis, prevention or treatment of diseases which  
CC involve the modified immunologically functional molecule, and agents  
CC which stimulate the activity of an immunologically functional molecule.  
CC The methods of the invention are used for the diagnosis, treatment and  
CC prevention of a broad range of diseases including cancer, circulatory  
CC disease, viral or bacterial infection, allergy, autoimmune disease and  
CC inflammation. The present sequence represents a rat FUR8 cDNA isolated

CC In an exemplification of the invention.  
XX  
SQ Sequence 979 Bp; 283 A; 223 C; 220 G; 253 T; 0 other;  
Qy  
Best Local Similarity 47.9%; Score 827; DB 21; Length 979;  
Best Local Similarity 90.3%; Pred. No. 2,9e-226;  
Matches 884; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
  
Qy 720 ACTGCGCTTGGATCTACAAATTTGGCGCTACCTACTGGGGAGTGGAAACGTGTGTTAG 779  
Db 1 ACTCATCTTGGATCTACAAATTTGGCGCTACTGCTACTGGATGGAGACTGTCTTTAG 60  
Qy 780 ACCTGTAGTGAAGCCTGACAGACAGATCTGACACCTCCACTGACATTTGTCAGGTGA 839  
Db 61 ACCTGTAGTGAAGCAGATGACAGACAGATCTGACCTCCACTGACATTTGTCAGGTGA 120  
Qy 840 ACTGAAGGACAAATTTGTCAGGTGTTGAGCTTCCCATTTTGTAGACAGTTTCATCTCG 899  
Db 121 AGTGAATGACAAATTTGTCAGGTGTTGAGCTTCCCATTTTGTAGACAGCTTCATCTCG 180  
Qy 900 TCCCTCATATTTACCCCTGCTGCCAGAGACCTTGGAGATCGACTTGTACGATCCGA 959  
Db 181 GCTCTCTTACTTACCACTGCTGCTTCCAGAGACCTTGCAGATCGACTGTAAGACTCCA 240  
Qy 960 TGGTGATCTCTGACAGTGTGTGCTATCCAGTTTGTCAAGTACTGATTCGCCACACCC 1019  
Db 241 TGGTGATCTCTGACAGTGTGTGCTATCCAGTTTGTCAAGTACTGATTCGCCACACCC 300  
Qy 1020 CTGGCTGGAAAGAAATAGAAAGAGCCACAGAAAGCTTAGGCTTCAAACTCCACTTAT 1079  
Db 301 TTGGCTGAAGAAAGAAATAGAAAGAGCCACAGAAAGCTTGAACATCCACTCAT 360  
Qy 1080 TGGAGTCCATGTTAGAGCGCAGACAGAAAGTGGAGGCGAAGCCCTTCATCCCATTTGA 1139  
Db 361 TGGAGTCCATGTTAGAGCGCAGACAGAAAGTGGAGAGAGGCGCTTCATCCCATTTGA 420  
Qy 1140 GGAATACACGCTCAGCTTGAAGAAGACTTTCAGCTTCTTCTGCGAGATGCAAGTGA 1199  
Db 421 AGAGTACATGCTACATGTTTGAAGAACATTTTTCAGCTTCTGCGAGATGCAAGTGA 480  
Qy 1200 TAAAAAAAGGCTATTTTGGCCACAGATGACCTGCTTTGTAAAGAGCAAAACAAA 1259  
Db 481 TAAAAAAAGGCTATTTTGGCCACAGATGACCTGCTTTGTAAAGAGCAAAACAAA 540  
Qy 1260 GTACCCAGTATGAAATTTTATGATATCTATCTCTGCTGACAGTACATATA 1319  
Db 541 GTACCCAGTATGAAATTTTATGATATCTATCTCTGCTGACAGTACATATA 600  
Qy 1320 TCGATATACAGAAATTTCACTTGGGGGTGATCTCGATATACACTTCTCTCCAGGC 1379  
Db 601 TCGATATACAGAAATTTCACTTGGGGGTGATCTCGATATACACTTCTCTCCAGGC 660  
Qy 1380 AGACTTCCTAGTGTGATCTTTCATCGCAGGCTGTAGAGTTCCTTATGAATCATGCA 1439  
Db 661 TGACTTCCTAGTGTGATCTTTCATCGCAGGCTGTAGAGTTCCTTATGAATCATGCA 720  
Qy 1440 AGCGTCGATCTGATAGCTCTGCAACTCCGTTTGGATGAGATCTACTATTTGG 1499  
Db 721 AACCTGCACTCTGATGCTCTGCAAACTTCCACTTTTATGATGACATCTACTATTTGG 780  
Qy 1500 AGGCCCAAAATGCCCCACACCAAAATTCGCAATTCCTCACCACCTCGAAGTGAAGAGA 1559  
Db 781 AGGCCCAAAATGCCCCACACCAAAATTCGCAATTCCTCACCACCTCGAAGTGAAGAGA 840  
Qy 1560 AATCCCCATGGAACCTGAGATATTTATGTGTGCTGGAAATCACTGGATGGCTATCC 1619  
Db 841 AATCCCCATGGAACCTGAGATATTCATTTGTGTGCTGGAAATCACTGGATGGCTATCC 900  
Qy 1620 TAAAGGTGTTAAAGAAACTGGAGAGAGGCGCTATATCCCTGACAAAGTTGAGAGA 1679  
Db 901 TAAAGGTGTTAAAGAAACTGGAGAGAGGCGCTATATCCCTGACAAAGTTGAGAGA 960  
Qy 1680 GAAGATAGAAACAGTCAG 1698

```
Db 961 GAAGATAGAAACGGTCAAG 979
|||||
RESULT 13
ABK70037
ID ABK70037 standard; cDNA; 979 BP.
XX
AC ABK70037;
XX
DT 15-JUL-2002 (first entry)
XX
DE Antibody production method related cDNA #5.
XX
KM Antibody production; cytostatic; immunomodulator; vasotropic; virucide;
KM antibacterial; antiinflammatory; antiallergic; allergy; inflammation;
KM autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;
KM tumor; circulatory disease; infection; primer; ss.
XX
OS Rattus norvegicus.
XX
PN WO200231140-A1.
XX
PD 18-APR-2002.
XX
PF 05-OCT-2001; 2001WO-JP08804.
XX
PR 06-OCT-2000; 2000JP-0308526.
XX
PA (KYOM ) KYOMA HARUKO KOGYO KK.
XX
PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
DR WPI: 2002-340182/37.
XX
PT Cells producing antibody compositions including antibody fragments and
PT fusion proteins with Fc domain of antibody, useful for prevention or
PT treatment of cancer, immune diseases, circulatory diseases and
PT infections -
XX
PS Example 9; Page 13-14; 314pp; Japanese.
XX
CC This invention relates to novel method for antibody production
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell
CC transfected with a gene encoding an antibody molecule for producing a
CC composition comprising an antibody molecule with an Fc domain bonded
CC to the N-glycoside linkage complex sugar chain. The produced antibody
CC compositions are drugs for prevention or treatment of diseases
CC accompanying tumor, allergy or inflammation, autoimmune diseases,
CC circulatory diseases, and viral and bacterial infections. The
CC antibodies can be stably produced using the method of the invention
CC with high binding activity and potency thus leading to high safety and
CC reduced side effects when applied alone or in combination with other
CC drugs for therapy. The present sequence represents a nucleotide
CC molecule used in the method of the invention.
XX
SQ Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;
Query Match 47.9%; Score 827; DB 24; Length 979;
Best Local Similarity 90.3%; Pred. No. 2,9e-226;
Matches 884; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
```

```
Db 121 AGTGAATGACAAAATATTCAGGTGGAGGCTCCCATTTTGAAGACGCTTCATCTTCG 180
QY 900 TCCTCATATTTTACCCCTGGCTGTCCAGACACCTTGCAGATTCGACTTGTACAGTCCA 959
Db 181 GCCTCCTTACTTACCACTGTGGCTGTCCAGACACCTTGCAGATTCGACTTGTAAAGTCCA 240
QY 960 TGGTGATCCTGCAGTGTGGGTATCCGAGTTTGTCAAGTACTTGATTTGGCCCAACC 1019
Db 241 TGGTGATCCTGCAGTGTGGGTATCCGAGTTTGTCAAGTACTTGATTTGGCCCAACC 300
QY 1020 CTGGCTGGAAGAAAGAAATGAGAGGCCACCAAGAGCTAGGCTTCAACATCCAGTTAT 1079
Db 301 TTGGCTGGAAGAAAGAAATGAGAGGCCACCAAGAGCTAGGCTTCAACATCCAGTTAT 360
QY 1080 TGGAGTCACTGTTAGACGACACAGACAAATGGAGCGAAGCAAGCCCTTCCATCCATTGA 1139
Db 361 TGGAGTCACTGTTAGACGACACAGACAAATGGAGCGAAGCGAAGCCCTTCCATCCATTGA 420
QY 1140 GGAATACAGGCTGCAGCTTGAAGAGCTTTCAGGCTTCTGCTCCGAGATGCAAGTGA 1199
Db 421 AGAGTACATGTATCATGTTGAAGAACATTTTCAAGCTTCTGCGACGAGATCCAGTGA 480
QY 1200 TAAAAAAGAGGTATATTTGGCCACAGATGACCTGCTTTGTTAAAAAGAGCAAAACAAA 1259
Db 481 TAAAAAAGAGGTATATTTGGCCACAGATGACCTGCTTTGTTAAAAAGAGCAAAACAAA 540
QY 1260 GTACCCCACTTATGAATTTATAGATATACCTATCTTGTGGTACGTGACATCATTA 1319
Db 541 GTACTCCAAATTTGAATTTATAGTATTAACCTATTTCTGTCAGCTGACATCAACAA 600
QY 1320 TCGATATACGAAATTTCACTTCGGGGGTGATCCGGAATATACATTTCTCTCCAGGC 1379
Db 601 TCGATATACGAAATTTCACTTCGGGGGTGATCCGGAATATACATTTCTCTCTCAGGC 660
QY 1380 AGACTTCCTAGTGTACTTTTTCATCGCAGGTGTGTAGAGTGTGTTATGAATCATGCA 1439
Db 661 TGACTTCCTAGTGTACTTTTTCATCGCAGGTGTGTAGAGTGTGTTATGAATCATGCA 720
QY 1440 AGCGCTGCATCTGATGCTCTGCGACATTCCTTTTGGATGATCATTTATTTGG 1499
Db 721 AACCCGCAATCTCGATGCTCTGCAAACTTCACATCTTTAGATGATCATTTATTTGG 780
QY 1500 AGGCCCAAAATGGCCACACAAATTTGCCATTTATCTCTCAACACCTGCAAGTGAAGAGA 1559
Db 781 AGGCCCAAAATGGCCACACAGATTTGCCATTTATCTCTCAACACCTGCAAGTGAAGAGA 840
QY 1560 AATCCCATGGAACCTGAGATATTTATGTTGGTGTGCTGGAATATCACTGGATGGCTATCC 1619
Db 841 AATCCCATGGAACCTGAGATATTTATGTTGGTGTGCTGGAATATCACTGGATGGTATTTTC 900
QY 1620 TAAAGGTGTTAACAGAAACCTGGGAAGACGGGCTTATATCCCTCTACAAAGTTGAGA 1679
Db 901 TAAAGGTGTTAACAGAAACCTGGGAAGACGGGCTTATATCCCTCTACAAAGTTGAGA 960
QY 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACGGTCAAG 979
RESULT 14
AAF87953
ID AAF87953 standard; DNA; 699 BP.
XX
AC AAF87953;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.
XX
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunosay;
KW ds.
```

```

OS Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1.639
FT /tag= a
FT /partial
FT /product= "alpha 1-6 fucosyltransferase 343-575"
PN JP2001011097-A.
PD 16-JAN-2001.
PE 29-JUN-1999. 99JP-0183569.
PR 29-JUN-1999. 99JP-0183569.
XX (FJRE ) FUJIREBIO KK.
PA WPI: 2001-275926/29.
DR P-PSDB: AAB75062.
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunosassay -
XX Example 1; Page 7-8; 11pp; Japanese.
PS
XX
XX The present invention describes an anti-human alpha 1-6
XX fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX transferase) antibody. Also described are: (1) a hydridoma producing
XX the above monoclonal antibody; (2) an immunosassay for detecting human
XX alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX and (3) a reagent used for the above immunosassay. The anti-human
XX alpha 1-6 fuct antibody can be used for immunosassay. The present
XX sequence encodes the human alpha 1-6 fuct protein of residues 343 to
XX 575, which is used in an example from the present invention.
SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 35.1%; Score 606.8; DB 22; Length 699;
Best Local Similarity 91.8%; Pred. No. 2.8e-163;
Matches 641; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 1027 GAAAGGAAATAGAGAGAGCCACCAAGAGCTAGGCTTCAACATCCAGTTATTGAGTC 1086
DB 1 GAAAGGAAATAGAGAGAGCCACCAAGAGCTAGGCTTCAACATCCAGTTATTGAGTC 60
OY 1087 CAGTGAAGCGCACAGCAAGAGTGGAGCGAGAGCCCTTCATCCATCCATGAGGAATAC 1146
DB 61 CAGTGAAGCGCACAGCAAGAGTGGAGAGAGCTGCTTCATCCATCCATGGAAGATAC 120
OY 1147 ACGGTGACAGTTGAGAGAGACTTTCAGCTCTGTCGAGAAATGCAAGTGATTAAGAAA 1206
DB 121 ATGGTGACATGTTAAGAACATTTTCAAGCTTCTTGACAGCGAGATGCAAGTGACAAA 180
OY 1207 AGGGTGATTTGGCCACAGATGACCTGCTTGTAAAGAGGCAAAACAAAGTACCCC 1266
DB 181 AGAGTGATTTGGCCACAGATGACCTGCTTGTAAAGAGGCAAAACAAAGTACCCC 240
OY 1267 AGTTATGAATTTTATGATATCTATCTCTGTCAGCTGAGCTGACTACTATATCATAT 1326
DB 241 AATTAATGAATTTTATGATATCTATCTCTGTCAGCTGAGCTGACTACTATATCATAT 300
OY 1327 ACAGAAATTTCACTTGGGGGTGATCTGATATACATTTCTCTCCAGGAGACTTC 1386
DB 301 ACAGAAATTTCACTTGGGGGTGATCTGATATACATTTCTCTCTCAGGAGACTTC 360
OY 1387 CTAGTGTGACTTTTTCATCGCAGGCTGTAGAGTGTCTTATGAATCATCAAGGCTTC 1446
DB 361 CTAGTGTGACTTTTTCATCGCAGGCTGTGAGATTCATTAATTAATGAACAACTA 420
OY 1447 CATCCGATGCGCTCTCGCAACTTCGCTTCTTGATGATGACATCTACTATTTTGAAGCCCA 1506
DB 421 CATCCGATGCGCTCTCGCAAACTTCATCTTTAGATGACATCTACTATTTTGGGGCCAG 480

```

```

OY 1507 AATGCCACAAACCAATTTGCTATTCCTACCAACTGACATGAGAGAAATTCGCC 1566
DB 481 AATGCCACAAATTTGCTATTTGCTACACCAACCCGAACTGACATGAAATTTCCC 540
OY 1567 ATGGAACCTGGAGATATTTATTTGGTGTGCTGGAATTAATCACTGGAGATGCTTAAAGCT 1626
DB 541 ATGGAACCTGGAGATATTTATTTGGTGTGCTGGAATTAATCACTGGAGATGCTTAAAGCT 600
OY 1627 GTTACAGAGAAACTGGGAGAGGAGGCTATATCCCTCCACAAAGTGGAGAGATTA 1686
DB 601 GTTACAGAGAAATTTGGAGAGGAGGCTATATCCCTCCACAAAGTGGAGAGATTA 660
OY 1687 GAAACAGTCAAGTATCCCAATATCCCAAGGCTGAGACA 1724
DB 661 GAAACGTCAGTATCCCAATATCCCAAGGCTGAGACA 698

RESULT 15
ABLO4601
XX ID ABLO4601 standard; cDNA; 2761 BP.
XX AC ABLO4601;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8285.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ss.
XX PN Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR P-PSDB: ABB60498.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 8285; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL57737-ABR72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 2761 BP; 657 A; 731 C; 755 G; 618 T; 0 other;

Query Match 20.3%; Score 351.6; DB 23; Length 2761;
Best Local Similarity 55.9%; Pred. No. 6.9e-90;
Matches 694; Conservative 0; Mismatches 539; Indels 9; Gaps 1;

```



QY	473	ATCATGAAAGGCTATTAATGAGGAGATCTTATACCTCACTGATCAACAGATGGGGCAGGTC	532
Db	1083	ACGACAGCGCTCCCTGCTGAGCGATATGAGCGGATGCGCTAGCTCCGATGGCTTACGAGG	1142
QY	533	ATTGGCGTGAAGAAGGAGGCGCAAGATCTGACAGAGCTGGGCGCAGGAAATTAACATATC	592
Db	1143	CGTGGGAGACACAAGAGGAGGCAJGSGGATTTGAGCGATTTGGTACAGCGGGCGCTCAGCACC	1202
QY	593	TTTCAGATCCCAAGAGCTGACGCAAGAACCAAGAGCTAGTGTAAATATCAACAAAGGCT	652
Db	1203	TACAGATCCCGAGTGTGTCAGAAATGCTCGCAACTGGCTGTAAAGCTCAACAAAGGCT	1262
QY	653	GTGGCTATGGCTGTGCTGCTCCATCATGTAGTGTAGTCTTATGATTTGCTATGAGCACC	712
Db	1263	GTGGCTATGGCTGTGCTGCTCCATCATGTAGTGTAGTCTTATGATTTGCTATGAGCACC	1322
QY	713	ACGCAACACTGCGCTTGGATATCACAATTTGGCGCTACGCTACTGGGGGATGGAAACTG	772
Db	1323	ACGCAACACTGCGCTTGGATATCACAATTTGGCGCTACGCTACTGGGGGATGGAAACTG	1382
QY	773	TGTTTACAGCTTGTAGTGTAGAGCTGACACAGATCTGGCAGCTCCACTGACATTTGGT	832
Db	1383	TGTTTACAGCTTGTAGTGTAGAGCTGACACAGATCTGGCAGCTCCACTGACATTTGGT	1442
QY	833	CAGCTGAAATTAAGGACAAATAATGTTTACAGTGGTTGAGCTCCCATTTGTAGACAGTTTC	892
Db	1443	CAGCTGAAATTAAGGACAAATAATGTTTACAGTGGTTGAGCTCCCATTTGTAGACAGTTTC	1493
QY	893	ATCTCTGCTCTCCATTTATTTACCCCTGGCTGTCCAGAAAGCTTTGACAGATGACTGTAC	952
Db	1494	TGCGGAGAGCGCGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1553
QY	953	GAGTCCATGCTATCTGCAATGTTGGTGGGTATCCAGTTTGTCAAGTACTTATTTGCGC	1012
Db	1554	GCTGATGAG	1613
QY	1013	CACAACTTGGCTGGGAAAGAAATGAAAGAGGCGCACAAAGAAAGTACGCTTCAAACTC	1072
Db	1614	CACAACTTGGCTGGGAAAGAAATGAAAGAGGCGCACAAAGAAAGTACGCTTCAAACTC	1673
QY	1073	CAGTTATGAGATGCTATGTTAGAGGACAGACAAAGTGGGAGGAGAGAGAGAGAGAGAG	1132
Db	1674	CAGTTATGAGATGCTATGTTAGAGGACAGACAAAGTGGGAGGAGAGAGAGAGAGAGAG	1733
QY	1133	CCATTGAGGAAATACAGGCTGACGTTGAAAGAACTTTCAGCTTCTGCTCGCAGAAATGC	1192
Db	1734	GGTGGAGAGATATATGAGCTATGTGAGAGAGCTACAGCGCAGCTGAGAGTAAAGCGCA	1793
QY	1193	AAGTGATTAATAAAGGAGGATTTTGGCCACAGATGACCCGCTTTGTAAAGAGGCAA	1252
Db	1794	GCAGTGTGCTGCGCGGATTTTCTGCGATCGGATGATGCCAGGTTATGAGAGAGCGC	1853
QY	1253	AAACAAGTACCCCAAGTTATTAATTAATGATTAATGATTAATGATTAATGATTAATGAT	1312
Db	1854	GGCGAAAGTACCCCAAGTTATTAATTAATGATTAATGATTAATGATTAATGATTAATGAT	1913
QY	1313	TACATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT	1372
Db	1914	TATTCACAGAGATACAG	1973
QY	1373	CCGAGGAGAGCTTCTGATGTTGATTTTTCATGCGAGGCTGTGAGAGTGTCTTATGAAA	1432
Db	1974	CCATGTCGATCATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2033
QY	1433	TGATGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1492
Db	2034	TATGCGAGAGATGATTCGAGATGAGCGAGCTGTTCAAGTGTGAGAGAGATTAAT	2093
QY	1493	ATTTTGAAGGAG	1552
Db	2094	ACTAGGAG	2153
QY	1553	AAGGAG	1612

Search completed: November 1, 2002, 21:39:33  
Job time : 286 secs

Search completed: November 1, 2002, 21:39:33  
Job time : 286 secs

Accession	Sequence	Position
Db	ACGAGAGTCTACAGCTCTCGTTGGTGACCTCTCTCGGTGGCTGTACCACTTGGATG	2213
Qy	GCTATCTTAAAGTGTTACGAAACGTGGAGAGCAGGGCCCTATATCCCTCTCAACAAG	1672
Db	GCAATTCGCAAGGGCAGCAATACCCGACCAACAGGAGGGCCGTATTCCTCGTTCAAGG	2273
Qy	TTTCGAGGAGATGAAACAGCTCAAGTACCCCAATATCCG	1714
Db	TGGAGAGAAAGGTGACACCCGACCAAGCTGCTCTCTATCCGG	2315

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 20:37:08 : Search time 3099 Seconds

(without alignments)  
16227.702 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728  
Sequence: 1 atgcggcactgactgcttc.....atcccgagctgacagtaaa 1728

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rpd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1726.4	99.9	1728	6	AR170074	AR170074 Sequence
2	1726.4	99.9	1728	6	E14720	E14720 porcine mRN
3	1726.4	99.9	1836	6	D86723	D86723 porcine mRN
4	1545.6	89.4	1728	4	AF247186	AF247186 Bos tauru
5	1510.4	87.4	2002	9	D89289	D89289 Homo sapien
6	1510.4	87.4	2100	6	E15725	E15725 Human mRN
7	1510.4	87.4	2898	9	HSY17976	HSY17976 Homo sapien
8	1510.4	87.4	2992	9	HSY17977	HSY17977 Homo sapien
9	1510.4	87.4	3186	9	HSY17978	HSY17978 Homo sapien
10	1510.4	87.4	3280	9	HSY17979	HSY17979 Homo sapien
11	1505.6	87.1	2100	6	AR170077	AR170077 Sequence
12	1467.8	84.9	2976	10	BC010666	BC010666 Mus muscu
13	1464.6	84.8	2176	10	AB025198	AB025198 Mus muscu
14	1215.4	70.3	2682	9	BC025385	BC025385 Homo sapi
15	950	55.0	1850	9	AF052088	AF052088 Homo sapi
16	878.4	50.8	1017	6	E43925	E43925 Antlbody re
17	750	43.4	1772	9	AB049740	AB049740 Homo sapi
18	606.8	35.1	699	6	E43926	E43926 Antlbody re
19	351.6	20.3	2764	3	AY051451	AY051451 Drosophil
20	276.8	16.0	1147	9	AB049828	AB049828 Homo sapi
21	276.8	15.7	187192	2	CNS018P0	AL108847 Human chr
22	270.8	15.7	187192	2	AC128532	AC128532 Rattus no
23	219.4	12.7	359	9	HSALSFUC1	AF038280 Homo sapi
24	215	12.4	34322	2	AC018043	AC018043 Drosophil
25	215	12.4	160038	2	AC023739	AC023739 Drosophil
26	213	12.3	366	9	AB03256756	AB032572 Homo sapi
27	185.4	10.7	148068	9	CNS018RN	AL161871 Human chr
28	183.8	10.6	662	9	AB03256752	AB032568 Homo sapi
29	175.8	10.2	168601	10	AC110302	AC110302 Mus muscu
30	155.4	9.0	335	9	HSALSFUC2	AF038281 Homo sapi
31	149.6	8.7	407	9	AB03256754	AB032570 Homo sapi
32	138.6	8.0	187192	2	AC128532	AC128532 Rattus no
33	136.6	7.9	351	9	AB03256757	AB032573 Homo sapi
34	129.2	7.5	221521	2	AC129041	AC129041 Rattus no
35	108.6	6.3	75282	2	DMBR37M9	AL133495 Drosophil
36	107.8	6.2	125	9	AB03256755	AB032571 Homo sapi
37	98	5.7	40172	3	AF022568	AB032578 Homo sapi
38	95.4	5.5	264	9	AB03256753	AB032568 Caenorhab
39	75.6	4.4	7218	6	I66494	I66494 Sequence 14
40	72.4	4.2	68336	2	AC100653	AC100653 Mus muscu
41	50	2.9	164304	2	AP002398	AP002398 Homo sapi
42	50	2.9	174569	2	AC011123	AC011123 Homo sapi
43	50	2.9	208532	2	AP001460	AP001460 Homo sapi
44	50	2.9	210839	2	AP001562	AP001562 Homo sapi
45	50	2.9	215374	2	AC090396	AC090396 Homo sapi

ALIGNMENTS

RESULT 1  
AR170074 LOCUS  
DEFINITION Sequence 1 from patent US 6291219.  
ACCESSION AR170074  
VERSION AR170074.1 GI:17908033  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1728)  
AUTHORS Taniguchi,N., Uozumi,N., Shiba,T. and Yanagida,I.S.  
TITLE .alpha.1-6 fucosyltransferase  
JOURNAL Patent: us 6291219-A 1 18-SEP-2001;  
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a



PI TANICUCHI NAOKUKI, UOZUMI HISAFUMI, SHIBA TETSUO PC  
C12N15/09.C07H21/04.C12N5/10.C12N9/10.(C12N5/10.C12R1.91). PC  
(C12N9/10).  
PC (C12R1.91):  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..1728  
FT /organism='Sus sp.'  
FT /tissue\_type='brain'  
FT CDS 1..1728  
FT /product='alpha-1,6-fucosyltransferase'.  
FEATURES  
source location/Qualifiers  
1..1728  
/organism='Sus sp.'  
/db\_xref='taxon:9826'  
BASE COUNT 521 a 362 c 419 g 426 t  
ORIGIN  
Query Match 99.9%; Score 1726.4; DB 6; Length 1728;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 ATCGGCGCATGAGCTGCTGCGGCGTGGATTATGCTCATTTCTTTGGCTGGGGACC 60  
DB 1 ATCGGCGCATGAGCTGCTGCGGCGTGGATTATGCTCATTTCTTTGGCTGGGGACC 60  
OY 61 TTGCTATTTTACATAGGTGGTCACTTGGTACGAGATATGACCACTCTGATCACTAGC 120  
DB 61 TTGCTATTTTACATAGGTGGTCACTTGGTACGAGATATGACCACTCTGATCACTAGC 120  
OY 121 CGGAAGACTGTCGAAGATTTTGGCAAAAGCTGGAAACCGTTAAACAACAATAATGAAGCTTG 180  
DB 121 CGGAAGACTGTCGAAGATTTTGGCAAAAGCTGGAAACCGTTAAACAACAATAATGAAGCTTG 180  
OY 181 AGAGAGATGGCTGATCTCTCCGATATACGAAGGCCCTTGTATCAGGGGCCAGCTTCA 240  
DB 181 AGAGAGATGGCTGATCTCTCCGATATACGAAGGCCCTTGTATCAGGGGCCAGCTTCA 240  
OY 241 GGAAGATTCGTCGTTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATAATTAT 300  
DB 241 GGAAGATTCGTCGTTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATAATTAT 300  
OY 301 AAGAAACAAACTAAATATGTCGAGGAGAGATCATGAATTCCTAAGGAGAGAGATTGAA 360  
DB 301 AAGAAACAAACTAAATATGTCGAGGAGAGATCATGAATTCCTAAGGAGAGAGATTGAA 360  
OY 361 AATGAGACTAAAGAGCTCTGTTTTTCTCAAAAGTGAAGTGAAGAAATTTAAAGAAATTTA 420  
DB 361 AATGAGACTAAAGAGCTCTGTTTTTCTCAAAAGTGAAGTGAAGAAATTTAAAGAAATTTA 420  
OY 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480  
DB 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480  
OY 481 AGGTCTATATGACGATCTATCTACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540  
DB 481 AGGTCTATATGACGATCTATCTACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540  
OY 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGGCGGAGATTAACATATCTTCAGAAAT 600  
DB 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGGCGGAGATTAACATATCTTCAGAAAT 600  
OY 601 CCCAAGAGACTGACGAAAGCCAAAGAGCTGATGTAATATCAAAAGGCTGTGGCTAT 660  
DB 601 CCCAAGAGACTGACGAAAGCCAAAGAGCTGATGTAATATCAAAAGGCTGTGGCTAT 660  
OY 661 GCGTGTACAGCTCATCATCTAGTGTACTGCTTATATGATTCATATGCGACCCAGCGACAA 720  
DB 661 GCGTGTACAGCTCATCATCTAGTGTACTGCTTATATGATTCATATGCGACCCAGCGACAA 720  
OY 721 CTGGCCTTGGCAATCTCAATTTGGCGCTAGCTAGTGGGGATGGGAAACTGTGTTTAGA 780

DB 721 CTGGCCTTGGCAATCTCAATTTGGCGCTAGCTAGTGGGGATGGGAAACTGTGTTTAGA 780  
OY 781 CCTGTAGTGAAGACGTCACAGACAGATCTGGCAGCTCCACTGGACATTTGGCTGAA 840  
DB 781 CCTGTAGTGAAGACGTCACAGACAGATCTGGCAGCTCCACTGGACATTTGGCTGAA 840  
OY 841 GTTAAAGACAAATAATGTTACAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900  
DB 841 GTTAAAGACAAATAATGTTACAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900  
OY 901 CTTCCATATTTTACCCCGGCTGTCCAGAAAGACTCTTGAGATGAGCTTTGACAGATTCAT 960  
DB 901 CTTCCATATTTTACCCCGGCTGTCCAGAAAGACTCTTGAGATGAGCTTTGACAGATTCAT 960  
OY 961 GGTGATCTGTCAGTGTGTCGATGATCCCATTTGTCAAGTACTTGTATGCGCCCAACACC 1020  
DB 961 GGTGATCTGTCAGTGTGTCGATGATCCCATTTGTCAAGTACTTGTATGCGCCCAACACC 1020  
OY 1021 TGGCTGGAAAGAAATATGAAAGAGCGCACCAAGACTAGCTTCAAAACATCCAGTTATT 1080  
DB 1021 TGGCTGGAAAGAAATATGAAAGAGCGCACCAAGACTAGCTTCAAAACATCCAGTTATT 1080  
OY 1081 GGAATCCATGTTTACGACGACAGCAAAAGTGGAGCGGAACGACCTTCCATCCATTGAG 1140  
DB 1081 GGAATCCATGTTTACGACGACAGCAAAAGTGGAGCGGAACGACCTTCCATCCATTGAG 1140  
OY 1141 GAATACACGGTGCAGCTTGAAGAAGCTTTCACCTTCTGTCGCAAGATGCAAGTGAT 1200  
DB 1141 GAATACACGGTGCAGCTTGAAGAAGCTTTCACCTTCTGTCGCAAGATGCAAGTGAT 1200  
OY 1201 AAAAAAAGGTGTATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGGCAAAACAAAG 1260  
DB 1201 AAAAAAAGGTGTATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGGCAAAACAAAG 1260  
OY 1261 TACCCAGTATGAATTTTATGATGAATACCTATCTCTGTCGACCTGACATCAATAT 1320  
DB 1261 TACCCAGTATGAATTTTATGATGAATACCTATCTCTGTCGACCTGACATCAATAT 1320  
OY 1321 CGATATACAGAAATTCACCTTGGGGGTGATCTGTGATATACACTTCTCTCCAGGCA 1380  
DB 1321 CGATATACAGAAATTCACCTTGGGGGTGATCTGTGATATACACTTCTCTCCAGGCA 1380  
OY 1381 GACTTCCTAGTGTACTTCTTTTCAATGTCGACGTCGTGAAGATTCCTATGAATCATGCA 1440  
DB 1381 GACTTCCTAGTGTACTTCTTTTCAATGTCGACGTCGTGAAGATTCCTATGAATCATGCA 1440  
OY 1441 GCGCTGATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTATTTTGA 1500  
DB 1441 GCGCTGATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTATTTTGA 1500  
OY 1501 GGCCCAATATGCCCAACAACCAATTTGCCATTTATCTCACCACTCGAAGAGGAA 1560  
DB 1501 GGCCCAATATGCCCAACAACCAATTTGCCATTTATCTCACCACTCGAAGAGGAA 1560  
OY 1561 ATCCCATGGAACCTGAGATATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1620  
DB 1561 ATCCCATGGAACCTGAGATATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1620  
OY 1621 AAGGTGTTTAAAGAAAGCTGGAAGAGCGGCTATATCTCTCTCAAAAGTTCCAGAG 1680  
DB 1621 AAGGTGTTTAAAGAAAGCTGGAAGAGCGGCTATATCTCTCTCAAAAGTTCCAGAG 1680  
OY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAGTAA 1728  
DB 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAGTAA 1728  
RESULT 3  
D86723 1836 bp mRNA linear MAM 07-FEB-1999  
LOCUS Porcine mRNA for N-acetyl-beta-D-glucosaminide  
DEFINITION alpha-1,6-fucosyltransferase, complete cds.

ACCESSION D86723  
VERSION D86723.1 GI:1752753  
KEYWORDS N-acetyl-beta-D-glucosaminide alpha-1,6-fucosyltransferase:  
GDP-L-Fuc.  
SOURCE Sus scrofa adult brain cDNA to mRNA.  
ORGANISM Sus scrofa  
REFERENCE 1 (sites)  
AUTHORS Uozumi, N., Yanagidani, S., Miyoshi, E., Ihara, Y., Sakuma, T.,  
Gao, C.X., Teshima, T., Fujii, S., Shiba, T. and Taniguchi, N.  
TITLE Purification and cDNA cloning of porcine brain  
GDP-L-Fuc:N-acetyl-beta-D-glucosaminide  
alpha1-->6fucosyltransferase  
J. Biol. Chem. 271 (44), 27810-27817 (1996)  
JOURNAL 97066976  
MEDLINE  
REFERENCE 2 (bases 1 to 1836)  
AUTHORS Naofumi, U., Shusaku, Y., Elji, M., Yoshito, I., Takahiko, S.,  
Cong-Xiao, G., Tadashi, T., Shigeru, F., Tetsuo, S. and Naoyuki, T.  
TITLE Purification and cDNA cloning of porcine brain  
GDP-L-Fuc:N-acetyl-b-D-glucosaminide:alpha-1-6fucosyltransferase  
Unpublished  
JOURNAL 3 (bases 1 to 1836)  
REFERENCE Uozumi, N.  
AUTHORS Direct Submission  
TITLE Submitted (29-JUL-1996) Naofumi Uozumi, Osaka University Medical  
Journal School, Department of Biochemistry: Yamadaoka 2-2, Suita, Osaka  
565, Japan (E-mail:selka@biochem.med.osaka-u.ac.jp,  
tel:06-879-3421, fax:06-879-3429)  
FEATURES  
source  
1. 1836  
Location/Qualifiers  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/lisse.type="brain"  
/dev\_stage="adult"  
57. 1784  
/note="GDP-L-Fuc"  
/product\_start=1  
/product="N-acetyl-beta-D-glucosaminide  
alpha-1,6-fucosyltransferase"  
/protein\_id="BA013157.1"  
/db\_xref="GI:1752754"  
/translation="MRPWTGSWRIMILIFAWGTLLFYIGHLVRDNDHSDHSRELS  
KILAKLERLQONEDLRMAESLRIPREPIDGPGSGVRALEEDPMKKEJENKK  
QKNGKDEHLRRIENGAKELFLESEKLKLNELNGELOSRADEFLLDGHNE  
RSIMTDLVYLSOTDAGDREKAEKDLTELVRRIYLNPDCKSAKKLVNINCKG  
GYGCOLHHVYCFMIAVYGNPTLAEASHMRYATGMPVPRVPSCTDRCSSSGH  
WSGEVKDKVQVVEPLIVDSVHPRPYRLVLAYPEDLADLVKRGDPAYWYSQPKY  
LIRPOWLEKEIEEATKLGFRHNPVIGVHVRITDKGAEAAHPIEEYVHVEEDFL  
LARRQVDKRRVYLAIDDPALKEAKTKPSYEFISDANSISAGSLHNRYSLENGV  
IIDIHFLSQADELVCTFSSQCVRAVEIEMQALHPDASAMFRSLDIIYYEGFPNANOI  
AIDPHOPTREGELPMEPDIDIGVAGNHMDGPKGVNRKIGRTGLVSYVVRKEIEVK  
VPTYPEADK"

BASE COUNT 560 a 381 c 443 g 452 t  
ORIGIN

Query Match 99.9%; Score 1726.4; DB 4; Length 1836;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCGGCATGACACTGGTCTGTCGGCTGGATTATGCTATCTCTTTTGGCTGGGGACC 60  
DB 57 ATCCGGCATGACACTGGTCTGTCGGCTGGATTATGCTATCTCTTTTGGCTGGGGACC 116  
QY 61 TTGCTATTTTACATAGTGTCTACCTGTGATGATTAATACACACTGTGATCACTAGC 120  
DB 117 TTGCTATTTTACATAGTGTCTACCTGTGATGATTAATACACACTGTGATCACTAGC 176  
QY 121 CGAGAACTGTCGAAGATTTTGGCAAGCTGGAAGCTTAAGCAACAAATGAAGACTTG 180  
DB 177 CGAGAACTGTCGAAGATTTTGGCAAGCTGGAAGCTTAAGCAACAAATGAAGACTTG 236  
QY 181 AGGAGATGCTGATCTCTCCGAATACGAGAGGCCCATTTGATCAGGGGCGAGCTTCA 240

DB 237 AGGAGATGCTGATCTCTCCGAATACGAGAGGCCCATTTGATCAGGGGCGAGCTTCA 296  
QY 241 GGAAGCTTGTCTCTTAGAAGCAATTTATGAAGGCCCAAGACAGATTGAAATTTAT 300  
DB 297 GGAAGCTTGTCTCTTAGAAGCAATTTATGAAGGCCCAAGACAGATTGAAATTTAT 356  
QY 301 AAGAACAACACTAAATATGATCCAGGAAGATCATGAATCTCTAAGAGAGATTGAA 360  
DB 357 AAGAACAACACTAAATATGATCCAGGAAGATCATGAATCTCTAAGAGAGATTGAA 416  
QY 361 AATGAGACTAAGAGCTGTGTTTTTCTACAAAGTGATTTGAAGAAATTAAGAAATTTA 420  
DB 417 AATGAGACTAAGAGCTGTGTTTTTCTACAAAGTGATTTGAAGAAATTAAGAAATTTA 476  
QY 421 GAAGAAATGACTCCAAACATGCAGATGAATTTCTATCAGATTGGGACATCATGAA 480  
DB 477 GAAGAAATGACTCCAAACATGCAGATGAATTTCTATCAGATTGGGACATCATGAA 536  
QY 481 AGGCTATATATGACGATCTATCTACTACCTCAGTCAACAGATGGGCGAGTTGGCGT 540  
DB 537 AGGCTATATATGACGATCTATCTACTACCTCAGTCAACAGATGGGCGAGTTGGCGT 596  
QY 541 GAAAGAGGCCCAAGATCTGACAGAGCTGTCCAGCGGAGAAATTAACATATCTTCAGAA 600  
DB 597 GAAAGAGGCCCAAGATCTGACAGAGCTGTCCAGCGGAGAAATTAACATATCTTCAGAA 656  
QY 601 CCAAGAGCTGACGCAAGAGCAAGAGCTGTGTAATTCAAAGAGCTGTGGCTAT 660  
DB 657 CCAAGAGCTGACGCAAGAGCAAGAGCTGTGTAATTCAAAGAGCTGTGGCTAT 716  
QY 661 GGCTGTACGCTCATCATGATGATGATCTGTTATGATTAATGACATGACCCAGCAACA 720  
DB 717 GGCTGTACGCTCATCATGATGATGATCTGTTATGATTAATGACATGACCCAGCAACA 776  
QY 721 CTGCGCTTGAATCTCAAAATTTGGCGCTACGTAAGTGGGGATGGGAAATGTTTGA 780  
DB 777 CTGCGCTTGAATCTCAAAATTTGGCGCTACGTAAGTGGGGATGGGAAATGTTTGA 836  
QY 781 CCGTAAAGTGAAGCGTGCACAGACAGATCTGGGAGCTCCAGTGGACATTTGGTACGGAA 840  
DB 837 CCGTAAAGTGAAGCGTGCACAGACAGATCTGGGAGCTCCAGTGGACATTTGGTACGGAA 896  
QY 841 GTAAGGACAAAATGTTTCAGGTGTGATGACCTCCCATTTGTAGACAGTGTTCATCTCGT 900  
DB 897 GTAAGGACAAAATGTTTCAGGTGTGATGACCTCCCATTTGTAGACAGTGTTCATCTCGT 956  
QY 901 CCTTCATATTTACCCCTGCTGTCCAGAAAGACCTTGCAGATGACTTGTAGAGATTCAT 960  
DB 957 CCTTCATATTTACCCCTGCTGTCCAGAAAGACCTTGCAGATGACTTGTAGAGATTCAT 1016  
QY 961 GGGATCTGTCAGTGGTGGGTATCCAGATTTGTCAAGTACTTGAATTCGCCCAACCC 1020  
DB 1017 GGGATCTGTCAGTGGTGGGTATCCAGATTTGTCAAGTACTTGAATTCGCCCAACCC 1076  
QY 1021 TGGCTGAAAGAAATATGAAGAGGCCACCAAGAGCTAGCTTCAAAATCAGTTATT 1080  
DB 1077 TGGCTGAAAGAAATATGAAGAGGCCACCAAGAGCTAGCTTCAAAATCAGTTATT 1136  
QY 1081 GGAATCATTTTAGACGACACAGACAAAGTGGAGCGGAACAGCTTCCATTCATTTGAG 1140  
DB 1137 GGAATCATTTTAGACGACACAGACAAAGTGGAGCGGAACAGCTTCCATTCATTTGAG 1196  
QY 1141 GAATPACAGCTGACAGCTTGAAGAAAGCTTTCAGCTTCTGCGCAAGATGCAAGTGAAT 1200  
DB 1197 GAATPACAGCTGACAGCTTGAAGAAAGCTTTCAGCTTCTGCGCAAGATGCAAGTGAAT 1256  
QY 1201 AAAAAAGGCTGATTTTGGCACAGATGACCTGCTTTGTTAAAGAGGCAAAACAAAG 1260  
DB 1257 AAAAAAGGCTGATTTTGGCACAGATGACCTGCTTTGTTAAAGAGGCAAAACAAAG 1316  
QY 1261 TACCCAGTTATGAATTTATAGTGAATCTATCTCTTGGTACGTGACGTACATTAAT 1320

Db 1317 TACCCAGTTATGAATTTATAGTATTAACCTATCTTGGTCAAGCTGACATATAAT 1376  
QY 1321 CATATACGAAATTTACTTGGGGTGTGATCCTGATATACATTTCTCCCGAGCA 1380  
Db 1377 CATATACGAAATTTACTTGGGGTGTGATCCTGATATACATTTCTCCCGAGCA 1436  
QY 1381 GACTTCTAGTGTACTTTTATCATCGAGCTGTGAGATTGCTTATGAATCATGCA 1440  
Db 1437 GACTTCTAGTGTACTTTTATCATCGAGCTGTGAGATTGCTTATGAATCATGCA 1496  
QY 1441 GCGTGTGATCCTGATGCTCTCGCAATCTTCTTGGATGACATCTATTTTGA 1500  
Db 1497 GCGTGTGATCCTGATGCTCTCGCAATCTTCTTGGATGACATCTATTTTGA 1556  
QY 1501 GCGCCAAATGGCCACCAACCAATGTCATTTATCCACCACTGCACTGAAGAGA 1560  
Db 1557 GCGCCAAATGGCCACCAACCAATGTCATTTATCCACCACTGCACTGAAGAGA 1616  
QY 1561 ATCCCATGGAAGCTGAGATATTTATGCTGCTGGAATACATGAGATGCTTCT 1620  
Db 1617 ATCCCATGGAAGCTGAGATATTTATGCTGCTGGAATACATGAGATGCTTCT 1676  
QY 1621 AAGGTGTTAACAGAAACTGGAGAGAGCGGCTATATCCCTCTTACAAAGTTGAGAG 1680  
Db 1677 AAGGTGTTAACAGAAACTGGAGAGAGCGGCTATATCCCTCTTACAAAGTTGAGAG 1736  
QY 1681 AAGATGAAGACAGTCAATACCCCATATCCCGAGGCTGACAAAGTAA 1728  
Db 1737 AAGATGAAGACAGTCAATACCCCATATCCCGAGGCTGACAAAGTAA 1784

RESULT 4  
LOCUS AF247186 1728 bp mRNA linear MAM 24-APR-2000  
DEFINITION Bos taurus 6-alpha-L-fucosyltransferase (FUT8) mRNA, complete cds.  
ACCESSION AF247186  
VERSION AF247186.1 GI:7638406

KEYWORDS  
SOURCE  
ORGANISM

Bos taurus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 1728)  
Javaud, C., Dupuy, F., Maftah, A., Michalski, J.C., Otiol, R.,  
Pelt, J.M. and Julien, R.

REFERENCE  
AUTHORS  
TITLE  
Ancestral exon organization of FUT8, the gene encoding the  
alpha6-fucosyltransferase reveals successive peptide domains which  
suggest a particular 3D core-structure for the  
alpha6-fucosyltransferase family  
unpublished

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
Submitted (20-MAR-2000) Unite de Genetique Moleculaire Animale,  
UR1061, Faculte des Sciences, 123, Avenue Albert Thomas, Limoges  
87060, France

FEATURES  
source  
Location/Qualifiers

1..1728  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/tissue\_type="lung"  
1..1728  
/gene="FUT8"  
1..1728  
/gene="FUT8"  
/product="6-alpha-L-fucosyltransferase"  
/protein\_id="AF65460.1"  
/db\_xref="gi:7638407"  
/translation="MRPWTGSMRWIMLILFAMGTLFLYIGSHLVARDNDHPHSHSRLS  
KILAKLERIKOONEDLRMAESLRIPEDIPDGPSGIRALEDLVAKKQIDENYK  
OTRNLGKIHIELRRRIENGAKELWFFLOSELKRLKLNENGLDHADEFLSDLGHNH

BASE COUNT 517 a 368 c 418 g 425 t  
ORIGIN  
Query Match 89.4%; Score 1545.6; DB 4; Length 1728;  
Best Local Similarity 93.4%; Pred. No. 0;  
Matches 1614; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
RSIMDLYLSOTDAGDMREKAEKDLTELVRRTTYLQNPDKDSKAKIYCNINKG  
GYGCOLHHVVCYPMFVAGTORLILIESNNRYATGMEVTPRVSFCTDGSYVTH  
MSGETKDNNOVVELPIYDSLHPRPPIPLAVPELDARLVRHGDPAVWVYQPVK  
LIRPOWLEKEIEEATKLGFRHPVIGYVTRTDVGEAFHPIEEYVIVEEHPOL  
LARMOVKRKYVLTADPSLKEAKTYVPIVEFISDSIMSAGLHRYENSLRGV  
ILDIHFLSOADFLVCTFSSOYCRVAVYEMOTLHPASNFSLDIDYFPGQAHNOI  
AIVHEPRTADEIPEMPEDDITGVAGNHMDYSKVNRLGRTGLPSYKVEKIEYK  
VPHVEAK"

Db 841 ATAAAGACAAAAATGTACAGTAGTCGAGCTCCCATTTGTGACAGATCTTCCTCGT 900  
QY CCTCATATTTTACCCGTGGCTGTCCAGAGACCTTGCAGATGAGCTGTAGAGTCAT 960  
Db 901 CCTCATATTTTACCCGTGGCTGTCCAGAGACCTTGCAGATGAGCTGTAGAGTCAT 960  
QY 961 GGTGATCTGTGAGTGTGGGTATCCAGATTTGTCAAGTATTTGTGCGCCACAACC 1020  
Db 961 GGTGATCTGTGAGTGTGGGTATCCAGATTTGTCAAGTATTTGTGCGCCACAACC 1020  
QY 961 GGTGATCTGTGAGTGTGGGTATCCAGATTTGTCAAGTATTTGTGCGCCACAACC 1020  
Db 961 GGTGATCTGTGAGTGTGGGTATCCAGATTTGTCAAGTATTTGTGCGCCACAACC 1020  
QY 1021 TGGCTGGAAGAAGAAATAGAGAGGCCACAGAGCTAGGCTTCAACATCCAGTTAT 1080  
Db 1021 TGGCTGGAAGAAGAAATAGAGAGGCCACAGAGCTAGGCTTCAACATCCAGTTAT 1080  
QY 1021 TGGCTGGAAGAAGAAATAGAGAGGCCACAGAGCTAGGCTTCAACATCCAGTTAT 1080  
Db 1021 TGGCTGGAAGAAGAAATAGAGAGGCCACAGAGCTAGGCTTCAACATCCAGTTAT 1080  
QY 1081 GGAGTCCATGTTAGAGCAGACAGAAAGTGGGCGGAACGACCTTCCATCCATTGAG 1140  
Db 1081 GGAGTCCATGTTAGAGCAGACAGAAAGTGGGCGGAACGACCTTCCATCCATTGAG 1140  
QY 1141 GAATACACGCTGACGCTTGAAGAAGACTTTCAGCTTCTTGCTGCGAGAAATGCAAGTGAT 1200  
Db 1141 GAATACACGCTGACGCTTGAAGAAGACTTTCAGCTTCTTGCTGCGAGAAATGCAAGTGAT 1200  
QY 1201 AAAAAAGGTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAG 1260  
Db 1201 AAAAAAGGTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAG 1260  
QY 1261 TACCCCATTTATATTTATTTAGTGAATCTATCTCTTGGTGCAGCTGACATACAAAT 1320  
Db 1261 TACCCCATTTATATTTATTTAGTGAATCTATCTCTTGGTGCAGCTGACATACAAAT 1320  
QY 1321 CGATATACAGAAATTCATCTTGGGGTGTGATCCTGGATATACACTTCTCTCCAGCA 1380  
Db 1321 CGATATACAGAAATTCATCTTGGGGTGTGATCCTGGATATACACTTCTCTCCAGCA 1380  
QY 1381 GACTTCTAGTGTGATCTTTTTCATGCGACAGCTGTAGAGTCTTATGAGAAATCAGCA 1440  
Db 1381 GACTTCTAGTGTGATCTTTTTCATGCGACAGCTGTAGAGTCTTATGAGAAATCAGCA 1440  
QY 1441 GCGCTCAGTCTGATGCTCTGCGAATCTCGTCTTGGATGACATCTCTATTTTGA 1500  
Db 1441 GCGCTCAGTCTGATGCTCTGCGAATCTCGTCTTGGATGACATCTCTATTTTGA 1500  
QY 1501 GGGCCAAATGCCACAAATTTGCTTATCTTACCACTGCACTGCAAGAGAA 1560  
Db 1501 GGGCCAAATGCCACAACTGATTTGCTTATCTTACCACTGCACTGCAAGAGAA 1560  
QY 1561 ATCCCATGAGACCTGAGATATTTTGGTGGCTGGAATACAGTGGATGCTATCCT 1620  
Db 1561 ATCCCATGAGACCTGAGATATTTTGGTGGCTGGAATACAGTGGATGCTATCCT 1620  
QY 1621 AAGAGTCTTAACAGAAAGTGGAGAGGAGGCTATATCTCTCAAGTTGCGAG 1680  
Db 1621 AAGAGTCTTAACAGAAAGTGGAGAGGAGGCTATATCTCTCAAGTTGCGAG 1680  
QY 1681 AAGATGAAGAACTGAGTACACCATATATCCGAGGCTGACAGATTA 1728  
Db 1681 AAGATGAAGAACTGAGTACACCATATATCCGAGGCTGACAGATTA 1728

RESULT 5  
Db8289  
LOCUS D89289 2002 bp mRNA linear PRI 07-FEB-1999  
DEFINITION Homo sapiens mRNA for N-Acetyl-beta-D-glucosaminide, complete cds.  
ACCESSION D89289.1 GI:2055306  
VERSION D89289.1 GI:2055306  
KEYWORDS N-Acetyl-beta-D-glucosaminide; GDP-L-Fuc; alpha 1-6  
SOURCE Fucosyltransferase: alpha1-6 FUCT.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (sites)  
Yanagidani,S., Uozumi,N., Ihara,Y., Miyoshi,E., Yamaguchi,N. and

Taniguchi,N.  
Purification and cDNA cloning of  
GDP-L-fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase  
(alpha1-6 Fuct) from human gastric cancer MKN45 cells  
J. Biochem. 121 (3), 626-632 (1997)  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Direct Submission  
Submitted (18-NOV-1996) Naoyuki Taniguchi, Osaka University Medical  
School, Department of Biochemistry, Yamadaoka 2-2, Suita, Osaka  
565, Japan (E-mail: prof.taniguchi@biochem.med.osaka-u.ac.jp,  
tel:81-6-879-3420, fax:81-6-879-3429)  
Location/Qualifiers  
1.2002  
/organism="Homo sapiens"  
/product="N-Acetyl-beta-D-glucosaminide"  
/protein\_id="BA19764.1"  
/db\_xref="taxon:9606"  
143.1870  
/note="GDP-L-Fuc; alpha 1-6 Fucosyltransferase; alpha1-6  
Fuct"  
/codon\_start=1  
/product="N-Acetyl-beta-D-glucosaminide"  
/translation="MRPWGSMRWIMLILFWAGTLEFYIGHLVNDNDIPDHSBELS  
KILAKLERKQONEDLRMAESLRPEIDGPAIGRVLEEDLVKAEQIDENYKK  
OTRNGKDEHLELRRIENGAKELMFLOSELKLNLENGELORLADDFLIDHNE  
RSIMDLYLSOTDGAOREKEMADLELVORITLYLONPKDSKAKKLVGNINIKG  
GYGCOLHNHYVCFMAYGTQRLILLESOMWRATGWEIVFPVSECTDRSGISTGH  
WSGEVKKNVQVLELPVDSLHPRPYLLAVLPEDLVLVHDPVAMWSQVRY  
LIRPQWLEKEIEATKLGFKHNPVIGVHVRDVGTEAAHPIEEVWVHEEHOL  
LARRMVDKRYVLATDPSLLEAKTKYRNEYISDINSISAGLHNRYTENSJRGV  
LIDHIFLSOADEFLVCTFSOVCRAVEIMOTLHPDASANFSLDLIDYVYGGONAHNIQ  
ALYAHOPRTADELIPMEPDILGVAGNHMGYSKGVNRKLGRTGLVPSYVREKLETVK  
YFTYPEAEK"

BASE COUNT 620 a 415 c 467 g 500 t  
ORIGIN  
Query Match 87.4%; Score 1510.4; DB 9; Length 2002;  
Best local Similarity 92.1%; Pred. No. 0;  
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
QY 1 ATGCGGCATGAGACTGGTTCGTGGCGTTGATTATGCTATCTTTTGGCTGGGAGC 60  
Db 143 ATGCGGCATGAGACTGGTTCGTGGCGTTGATTATGCTATCTTTTGGCTGGGAGC 202  
QY 61 TTGCTATTTTACATAGTGTGCTACCTGTTGACGAGATATATACCACTGTACTAGC 120  
Db 203 TTGCTATTTTATAGTGTGCTACCTGTTGACGAGATATATACCACTGTACTAGC 262  
QY 121 CGAAGACTGTCCAAAGATTTTGGCAAGAGCTGGAACGCTTAAACAAACAAATTAAGAACTTG 180  
Db 263 CGAAGACTGTCCAAAGATTTTGGCAAGAGCTGGAACGCTTAAACAAACAAATTAAGAACTTG 322  
QY 181 AGGAGATGCTGATCTCTCCGATATACGAGAGGCCCATTTGATCAGGGGACGCTTCA 240  
Db 323 AGGAGATGCTGATCTCTCCGATATACGAGAGGCCCATTTGATCAGGGGACGCTTCA 382  
QY 241 GGAAGATTTGCTCTTTAAGAAGCAATTTATGAAGCCAAAGACAGATTTGAAATTTAT 300  
Db 383 GGAAGATTTGCTCTTTAAGAAGCAAGCTTTAAGGCAAAACAGATTTGAAATTTAC 442  
QY 301 AAGAAACAACTAAATATGTTCCAGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 360  
Db 443 AAGAAACAGACCAAGAAATGTTCTGGGAGAGATCATGAATCTTGAAGAGAGAGATTGAA 502



[illegible]

Oy	121	CGAAGCTGTCGAAGATTTTGGCAAGCGTGGAAAGCGTTTAAACCAACAAATGAAAGACTTG	180
Db	318	CGAAGACTGTCCAAAGATTTGGGCAAAAGCTTGAACGCTTAAACACACAGATGAAGACTTG	377
Oy	181	AGGAAATGGCTGGATCTCTCCGAATACCAAGAAAGGCCCATTTGATCAGGGGCGCACTTCA	240
Db	378	AGGGCAATGGGCGCAATCTCTCCGGATTACCAAGAAAGGCGCATTTGATCAGGGGCGCACTTA	437
Oy	241	GGAGACTTGGTGGTTTAAAGAAAGCAATTTATGAGGGCCAAAGAACAGATTGAATTTAT	300
Db	438	GGAAAGATGACCGCTTTTAAAGAAAGCAGCTTGTAAAGGCCCAAGAAACAGATTGAATTTATC	497
Oy	301	AAGAACAACACTAAAAATGCTCCAGGAAAGAGATCATGAATATCTTAAGGAGCAGATTGAA	360
Db	498	AAGAACAAGACAGAAATGGTCTGGGGGAAGGATCATGAATCTCTGAGGAGAGATTGAA	557
Oy	361	AATGAGCTTAAAGAGCTCTGGTTTTTTCTTCAAAAGTGAAGTGAAGTAATTAAGAAATTA	420
Db	558	AATGAGACTTAAAGAGCTCTGGTTTTTTCTTCAAGAGTGAATTAAGAAATTAAGAAACTTA	617
Oy	421	GAAAGAAATGAACCTCCAAACACATGCAAGATGAAATTTCTATCAGATTTGGGACATCATGA	480
Db	618	GAAAGAAATGAACCTCCAAACACATGCAAGATGAAATTTCTTTGGATTTTGAACATCATGA	677
Oy	481	AGGCTCTAATGACGGATCTTACTACTCCTGATCAAAACAGATGGGGCAGGTGATGGCGT	540
Db	678	AGGCTCTAATGACGGATCTTACTACTCCTGATCAACAGATGGGAGCATGATGGCGG	737
Oy	541	GAAAGAGCGCCAAAGATCTGACAGAGCTGTGCCAGGAGAAATTAACATCTTCAGAT	600
Db	738	GAAAGAGCGCCAAAGATCTGACAGAGCTGTGGTCCAGGAGAAATTAACATCTTCAGAT	797
Oy	601	CCCAAGAGACTGACAGCAAGCGCAAGAGCTAGTGTGAATATCAACAAAGCGCTGGCTAT	660
Db	798	CCCAAGAGACTGACAGCAAGCGCAAGAGCTGTGTGAATATCAACAAAGCGCTGGCTAT	857
Oy	661	GGCTGTCAAGCTCATCTGTATGTACTGCTTTATGATTTGCATATGCGACCCAGCGAACA	720
Db	858	GGCTGTCAAGCTCATCTGTGTGTCTACTGCTTCATGATTTGCATATGCGACCCAGCGAACA	917
Oy	721	CTCGCTTTGGAAATCTCAATTTGGCGTCAGCGTACGCGGGGGATGGGAAATCTGTTTTGA	780
Db	918	CTCATCTTTGGAAATCTCAGAATTTGGCGCTATGCTACTGTGTGGATGGGAGACTGTATTAGG	977
Oy	781	CCTGTAAGTGAAGCGTGCACAGACAGATCTGGCAGCTCCACATGACATTTGGTCAGGTAA	840
Db	978	CCTGTAAGTGAAGCTGCACAGACAGATCTGGCATCTCCATGAGACACTGGTCAGGTAA	1037
Oy	841	GTAAGGACAAAATGTTTCAGAGGTGTGAGCTCCCATTTGTAGACAGTTCATCTCCGT	900
Db	1038	GTAAGGACAAAATGTTTCAGAGGTGTGAGCTCCCATTTGTAGACAGTTCATCTCCGT	1097
Oy	901	CCCTCATATTTACCCCTGGCGCTGTCCCAAGAAAGACTTGCAGATGCACTGTATAGAGTCCAT	960
Db	1098	CCCTCATATTTACCCCTTGGCTGTACCAAGAAAGACTTGCAGATGCACTGTATAGAGTCCAT	1157
Oy	961	GCTGATCTGCAAGTGTGTGGGTATCCCAAGTGTGTCAAGTACTTGAATTTGCCACACACC	1020
Db	1158	GCTGATCCCTGCAAGTGTGTGGGTCTCAAGTGTGTCAATTTCAATCCGCCACACAGCT	1217
Oy	1021	TGGCTGGAAGAAAGAAATAGAAGAGGCGCACCAAGAACTTGGCTTCAAAACATCAGTATT	1080
Db	1218	TGGCTAGAAAAAGAAATAGAAGAGGCGCACCAAGAACTTGGCTTCAAAACATCAGTATT	1277
Oy	1081	GGAGTCCATGTTAGACGCACAGACAAAGTGGAGCGGAAGCAGCTTCCATCCATTGAG	1140
Db	1278	GGAGTCCATGTTAGACGCCACAGACAAAGTGGAGCAGAGCTTCCATCCATTGAA	1337
Oy	1141	GAATATACGGTGCACGTTTGAAGAAAGCTTTGACCTTCTGTCTGCGAGAAATGCAAGTGCAT	1200
Db	1338	GAGTATATGTCATGTTTGAAGAAATTTTTCAGCTTCTTGGCAGCGAGATGCAAGTGCAT	1397

OY	1201	AAAAAAGGSGTATTTGGCCACAGATGACCTGGTTGTTTAAAGAGGCAAAACCAAG	1260
OY	1201	AAAAAAAAAGGTTATTTGGCCACAGATGACCTGGTTGTTTAAAGAGGCAAAACCAAG	1260
Db	1398	AAAAAAGAGTCTATTTGGCCACAGATGACCTGTTTATTTAAAGAGGCAAAACCAAG	1457
OY	1261	TACCCAGCTATGAAATTTATTAAGTAACTCATCTGTGGTCAGTGGACATCATAT	1320
Db	1458	TACCCCAATTTATGAAATTTATTAAGTAACTCATCTGTGGTCAGTGGACATCATAT	1517
OY	1321	CGATTTACAGAAATATTCACCTGGGGGTGGATCCGGAATATPACCTTTCTGCCAGCA	1380
Db	1518	CGATCCACAGAAATATTCACCTGGGGGTGGATCCGGAATATPACCTTTCTCTAGGCA	1577
OY	1381	GACTTCCTAGTGTGACTTTTTCATCGCAGGCTGTAGAGTTGCTTTATGAATCATGCA	1440
Db	1578	GACTTCCTAGTGTGACTTTTTCATCGCAGGCTGTGCGAGTTGCTTATGAATTTATGCA	1637
OY	1441	GCGCTGCATCTCGATGAGCCTCGCCGAAATTCCTGTTGGATGACATCATATTTTGA	1500
Db	1638	ACACTACATCTCGATGAGCCTCGCAACATCTTCATTTAGACATCTACTATTTTGGG	1697
OY	1501	GGCCCAATGCCCACACCAATTTGCCATTTATCTCCACCAACCTCGAATGAAAGCA	1560
Db	1698	GCCCAAGATGGCCCAACATCAATTTGCCATTTATGCTCACCAACCCCGAATCGCATGTA	1757
OY	1561	ATCCCCATGGAACCTGAGATATTTATTTGGTGTGGCTGGAAATCACATGGGATGCT	1620
Db	1758	ATTTCCCATGGAACCTGAGATATTCATTTGGTGTGGCTGGAAATCATTTGGATGGCTATCT	1817
OY	1621	AAAGGTGTTTACAGAAAACTGGGAAGAGAGGAGCTATATCCCTCTACAAAATTTGAGAG	1680
Db	1818	AAAGGTGTCAACAGGAAATTTGGGAAGAGAGGAGCTATATCCCTCTACAAAATTTGAGAG	1877
OY	1681	AAGATGAAACAGTCAAGTACCCCAATATCCCGAGGCTGACACAGTAA	1728
Db	1878	AAGATGAAACAGTCAAGTACCCCAATATCTCTGAGGCTGAGAAAAATTA	1925

RESULT 7	
LOCUS	HSX17976
DEFINITION	HSX17976 Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase, transcript B1.
ACCESSION	Y17976
VERSION	Y17976.1 GI:3451262
KEYWORDS	FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 Calleau A., Balanzino L., Candelier J.J., Oriol R. and Mollicone R.
TITLE	Differential splice variants of human FUT8 embryonic CDNA
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2898)
AUTHORS	Calleau A.L.V.
TITLE	Direct Submission
JOURNAL	Submitted (19-AUG-1998) A.L.V. Calleau, INSERM U178, 16 av P.V. Couturier, 94807 Villejuif Cedex, FRANCE
COMMENT	Related entry: D89289.
FEATURES	Location/Qualifiers
source	1..2898
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="14"
	/map="14q23"
	/clone="FUT8 transcript B1"
	/tissue_type="Whole embryo"
	/dev_stage="40-70 days"
gene	1..2874
	/gene="FUT8"
exon	1..101
	/gene="FUT8"

```
/note="B"
102..200
/ gene="FUT8"
/ note="C"
201..2156
/ gene="FUT8"
/ note="D"
429..2156
/ gene="FUT8"
/ EC_number="2.4.1.68"
/ note="1"
/ codon_start=1
/ product="glycoprotein 6-alpha-L-fucosyltransferase"
/ protein_id="CAA76985.1"
/ db_xref="GI:3451263"
/ db_xref="SPTREMBL:O00235"
/ translation="MRPTGSMRWITMLIFAWGTLFFYIGGLVLDNDHPDHSNELS
KILAKLERIKQONEDLRMAESLRIPEDPIDGPAIGRVYLEQVLAKEDIEYKK
OTRNGLGDHEILRRRIENGAKELFLOESLKLKLENGEJLORHAEFLDLHDHE
RSIMDLVYLSOTDGDMDREKAKDLIELVORRTYVLPNPKDSKARKKYNINIKGC
GYGCOLHHVYCFMIAVSTORTLILESQNMRYATGMEVTRPVSEETCDMSISTGH
WSGEVKDNVOVVELPIVDSLHPRPYLPPLAVBEDLADLVKHDPAVMVWSQFVKY
LIRQPMLEKEIEEATKRGFKHPYIGVHRTDKVGEAEHPLEEMVHVEHOL
LARMOVDKKRYLATDDPSLKEAKTYPNVEFSDNSISMSAGLHNRYSNLSGV
LIDHFLSQADELVCTFSSOYCRVAYEIMOTLHPDASANFSLDDIYFEGGNHNOI
ALYAHOPRTADEIPMEPGDIIIGVAGNHNDGYSKGVNRRKLGTLPSYKVRKIEYK
VPTPEAEK"
2218
/ gene="FUT8"
/ note="1"
/ frequency=".5"
/ replacement="c"
2758..2763
/ gene="FUT8"
/ note="putative"
2770..2775
/ gene="FUT8"
/ note="putative"
2869..2874
/ gene="FUT8"
/ note="putative"
BASE COUNT 881 a 589 c 641 g 787 t
ORIGIN
Query Match 87.4% Score 1510.4; DB 9; Length 2898;
Best Local Similarity 92.1% Pred. No. 0;
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
1 ATGCGGCATGAGACTGGTTCGTGGCGTTGGATTATGCTCATTTCTTTGCTGGGGACC 60
|||||
429 ATGCGGCCATGGACTGGTTCGTGGCGTTGGATTATGCTCATTTCTTTGCTGGGGACC 488
|||||
61 TTGCTATTTTACATAGTGGTCTGCTGTAGACGATTAATGACCACTGTATCTAGC 120
|||||
489 TTGCTGTTTATATAGTGGTCTGCTGTAGACGATTAATGACCACTGTATCTAGC 548
|||||
121 CGAGAACTGTCCAAAGTTTGGCAAGCTGGAACGCTTAAACACAAATGAAGCTTG 180
|||||
549 CGAGAACTGTCCAAAGTTTGGCAAGCTTGAACGCTTAAACACGAAATGAAGCTTG 608
|||||
181 AGGAGAACTGTGATCTCTCCGAATACAGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240
|||||
609 AGGCAATGCGCAATCTCTCCGATACAGAAAGGCCCATTTGATCAGGGCCAGCTTAA 668
|||||
241 GGAAGACTTCGTGCTTTAGAGCAATTTATGAAGGCCCAAGAACAGATTGAAAATTAT 300
|||||
669 GGAAGACTTCGTGCTTTAGAGCAATTTATGAAGGCCCAAGAACAGATTGAAAATTAC 728
|||||
301 AAGAAACAACATAAATGCTCCAGGAAGATCATGAATCTTAAGAGAGAGATTGAA 360
|||||
729 AAGAAACAACATAAATGCTCCAGGAAGATCATGAATCTTAAGAGAGAGATTGAA 788
|||||
361 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTA 420
|||||
```

```
|||||
Db 789 AATGAGCTAAGAGCTCTGTTTTTCTACAGAGGATTTGAAGAAATTAAGAACTTAA 848
|||||
421 GAAGAAATGAACCTCCMAAGACATGAGATGAATTTCTACAGATTTGGAGATCATGAA 480
|||||
849 GAAGAAATGAACCTCCMAAGACATGAGATGAATTTCTTGGATTTAGACATCATGAA 908
|||||
481 AGGCTATATAGAGATCTTATCTACCTGACGTCAAACAGATGGGACAGTATGGCGT 540
|||||
909 AGGCTATATAGAGATCTTATCTACCTGACGTCAAACAGATGGGACAGTATGGCGG 968
|||||
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGGAGAAATTAATCATATCTTCAGAT 600
|||||
969 GAAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGGAGAAATTAATCATATCTTCAGAT 1028
|||||
601 CCCAAGAGCTGACAGCAAAAGCCAAAGAGCTAGTGTGAATATCAACAAAGCTGTGCTAT 660
|||||
1029 CCCAAGAGCTGACAGCAAAAGCCAAAGAGCTGTGTGAATATCAACAAAGCTGTGCTAT 1088
|||||
661 GGCTGTACAGCTCCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
|||||
1089 GGCTGTACAGCTCCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148
|||||
721 CTCGCTTGAATCTGCAATTTGAGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 780
|||||
1149 CTCATCTTGAATCTGCAATTTGAGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 1208
|||||
781 CCTGTAGTGAAGCTGACAGACAGATCTGGACGCTGACAGCTGACAGCTGACAGCTGACAG 840
|||||
1209 CCTGTAGTGAAGCTGACAGACAGATCTGGACGCTGACAGCTGACAGCTGACAGCTGACAG 1268
|||||
841 GTTAAAGCAAAAATTTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
|||||
1269 GTTAAAGCAAAAATTTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1328
|||||
901 CCTCATATTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATGATGATGATGATGATGATGATGAT 960
|||||
1329 CCTCATATTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATGATGATGATGATGATGATGATGAT 1388
|||||
961 GGTGATCTGCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020
|||||
1389 GGTGATCTGCAAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1448
|||||
1021 TGGCTGGAATAAGAAATTAAGAGAGCCACCAAGAACTTACGCTTCAAAATTCACATTAT 1080
|||||
1449 TGGCTGGAATAAGAAATTAAGAGAGCCACCAAGAACTTACGCTTCAAAATTCACATTAT 1508
|||||
1081 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGAGCGGAAGCAAGCTTCCATCCATTTGAG 1140
|||||
1509 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGAGCGGAAGCAAGCTTCCATCCATTTGAG 1568
|||||
1141 GAATACAGCTGACAGCTTGAAGAAGCTTTCAGCTTCTTCTGCGCAAGATGAGAT 1200
|||||
1569 GAATACAGCTGACAGCTTGAAGAAGCTTTCAGCTTCTTCTGCGCAAGATGAGAT 1628
|||||
1201 AAAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTGTTAAAGAGGAGCAAAAACAAAG 1260
|||||
1629 AAAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTGTTAAAGAGGAGCAAAAACAAAG 1688
|||||
1261 TACCCAGTTTAAAGATTTTATGATGATTAACCTGATCTTGGTACAGCTGACATCAAT 1320
|||||
1689 TACCCAGTTTAAAGATTTTATGATGATTAACCTGATCTTGGTACAGCTGACATCAAT 1748
|||||
1321 CGATATACAGAAATTCACCTTGGGGGTGTGATCTGGATATACATCTTCTCCAGGCA 1380
|||||
1749 CGATATACAGAAATTCACCTTGGGGGTGTGATCTGGATATACATCTTCTCCAGGCA 1808
|||||
1381 GACTTCCTAGTGTGCTTTTTCATGCGAGGTCTGTAAGTGGTCTTAAGAAATCATGCA 1440
|||||
1809 GACTTCCTAGTGTGCTTTTTCATGCGAGGTCTGTAAGTGGTCTTAAGAAATCATGCA 1868
|||||
1441 GCGCTCATCTGATGCTCTGCGAAGCTTCCGTTCTTGGATGACATCTACTATTGGA 1500
|||||
```

Db 1869 ACACATACATCGTATGCTCTGCAAACTTCCATCTTCTTATAGATGACATCTACTATTTGGG 1928  
Qy 1501 GGCCCAATGCCACAAACCAATTTGCATTTATCTCACCACCTGCAACTGAGAGAA 1560  
Db 1929 GGGCAGATGCCACATCAATTCGATTTATGCTCACCACCCGAACTGCAGATGAA 1988  
Qy 1561 ATCCCATGGAACCTGAGATATTTATGCTGCTGGAATCATCGATGGCTATCTCT 1620  
Db 1989 ATTCCCATGGAACCTGAGATATCTGCTGCTGGAATCATCGATGGCTATCTCT 2048  
Qy 1621 AAAGGTCTTAACGAAACAGTGGAGAGAGGCGCTATATCCCTCTCAAAAGTTCCAGAG 1680  
Db 2049 AAAGGTCTTAACGAAATTTGGAAGAGAGGCGCTATATCCCTCTCAAAAGTTCCAGAG 2108  
Qy 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGCAAGTAA 1728  
Db 2109 AAGATGAAACGCTCAAGTACCCACATATCTCTGAGGCTGCAAGTAA 2156  
RESULT 8  
HSY17977  
LOCUS HSY17977 2992 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase,  
transcript B2.  
ACCESSION Y17977  
VERSION Y17977.1 GI:3451264  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2992)  
AUTHORS Calileau,A., Balanzino,L., Candelier,J.J., Oriol,R. and  
Mollicone,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2992)  
AUTHORS Calileau,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Calileau, INSERM U178, 16 av P.V.  
Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
location/Qualifiers  
source 1..2992  
/organism="Homo sapiens"  
/isolate="transcript B2"  
/db\_xref="taxon:9606"  
/map="14q23"  
/chromosome="14"  
/issue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..2874  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="B"  
102..200  
/gene="FUT8"  
/note="C"  
201..>2156  
/gene="FUT8"  
/note="D"  
429..2156  
/gene="FUT8"  
/BC\_number="2.4.1.68"  
/code="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76986.1"  
/db\_xref="GI:3451265"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPWTGSWRIMILIFAWGTLLEYIGHLVRDNHDPDHSRELS  
KILAKLERLKQONEDLRMAESLRIPGPIDGPALGRVRLLEEDLVKAKEOIENYKK

OTRNLGKDHDEILRRRIENGAKELMFLOSELKLRKLNLEGNELQRIHADEFLLDLCIHE  
RSIMTDLVYLSQDGAQWREKREKADDELYVORRITVLONPDCSKARKLYNJKGC  
GSCODLHVYVCFMIAVGTORTLIESOMRYATGMEVFPVPSGETCDRGI STGH  
WSGEKDKNOYVELPIYDLSLRPPRYLPLAYPELDARLRYVHDPPAMVMSOPVKY  
LIRPOWLEKEIEEATRKIGFKNPIYGVHVRTDYGTBAHPLEEVHYEHPOL  
LARRNOVDKRRVLYATDDPSLKEAKTKTPNPEFSDNISASGLHNHYTENS LGV  
LIDIHPLSQADELVCTFSQVCRVAVELMOTLHPASAFNHLSDIYVFGGONAIHQI  
ALYAHQPRTADEIPMEPGDIIGVAGNHMDYSKGVNRKLGIRGLYPYKVRKEIETVK  
YPTYPEAEK"  
2218  
variation  
/gene="FUT8"  
/note="11"  
/frequency=".5"  
/replace="C"  
2758..2763  
/gene="FUT8"  
/note="putative"  
2770..2775  
/gene="FUT8"  
/note="putative"  
2869..2874  
/gene="FUT8"  
/note="putative"  
polyA\_signal  
polyA\_signal  
polyA\_signal  
BASE COUNT 905 a 597 c 663 g 827 t  
ORIGIN  
Query Match 87.4%; Score 1510.4; DB 9; Length 2992;  
Best Local Similarity 92.1%; Pred. No. 0;  
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
Qy 1 ATCGGGCCATGAGACTGGTTCGTCGGCGTTGGATTATGCTATCTTTTGGCTGGGAGC 60  
Db 429 ATCGGGCCATGAGACTGGTTCGTCGGCGTTGGATTATGCTATCTTTTGGCTGGGAGC 488  
Qy 61 TTGCTATTTTACATAGCTGCTCACTTGCTGACGAGATATGACCACTGTACTCTAGC 120  
Db 489 TTGCTATTTTATATAGCTGCTCACTTGCTGACGAGATATGACCACTGTACTCTAGC 548  
Qy 121 CGAAGACTGTCGAAGATTTTGGCAAGCTGGAAGCGCTTAACACAAATGAAGACTTG 180  
Db 549 CGAAGACTGTCGAAGATTTTGGCAAGCTGGAAGCGCTTAACACACAGATGAAGACTTG 608  
Qy 181 AGGAGATGCTGATCTCTCCGAATACGAAAGGCCCATTTGATGAGGGCCAGCTTCA 240  
Db 609 AGGAGATGCTGATCTCTCCGAATACGAAAGGCCCATTTGATGAGGGCCAGCTTCA 668  
Qy 241 GGAAGATTGCTCTTGAAGAGCAATTTATGAAGGCCAAAGACAGATTTGAATTTAT 300  
Db 669 GGAAGATTGCTCTTGAAGAGCAATTTATGAAGGCCAAAGACAGATTTGAATTTAT 728  
Qy 301 AAGCAACAACTAAATATGCTCCAGGAAGAGATCATGAATCTTAAGAGAGAGATTTGAA 360  
Db 729 AAGCAACAGACGAATATGCTCCAGGAAGAGATCATGAATCTTAAGAGAGAGATTTGAA 788  
Qy 361 AATGAGCTAAAGAGCTGTGTTTTTTCTAACAAGTGAGTTGAAGAAATTTAAAGATTTA 420  
Db 789 AATGAGCTAAAGAGCTGTGTTTTTTCTAACAAGTGAGTTGAAGAAATTTAAAGATTTA 848  
Qy 421 GAAGGAATTAACCTCAAGACATGCAAGATTTTCTATCAATTTGGACATCATGAA 480  
Db 849 GAAGGAATTAACCTCAAGACATGCAAGATTTTCTATCAATTTGGACATCATGAA 908  
Qy 481 AGGCTATATGACGGATCTATCTACTAGTCAAAACAGATGGGCGAGTGTGGCGT 540  
Db 909 AGGCTATATGACGGATCTATCTACTAGTCAAAACAGATGGGCGAGTGTGGCGG 968  
Qy 541 GAAAGAGGCCAAAGATCTGACAGAGCTGTCACGGAGAAATTAATTTTCAGAAAT 600  
Db 969 GAAAGAGGCCAAAGATCTGACAGAGCTGTCACGGAGAAATTAATTTTCAGAAAT 1028  
Qy 601 CCCAAGAGCTGACGACAAAGCAAGACTAGTGTATATTCACAAAGGCTGTGCTAT 660  
Db 1029 CCCAAGAGCTGACGACAAAGCAAGACTAGTGTATATTCACAAAGGCTGTGCTAT 1088

QY 661 GGCTGTCACATCATCATAGTGTACTGCTTTATGATTCATATGACACCCAGCAGACA 720  
|||||  
DB 1089 GGCTGTGACCTCATCATATGTTGGTCTACTCTTCATATGATTCATATGACACCCAGCAGACA 1148  
QY 721 CTGCGCTTGGAAATCTCACAATTTGGCGCTACGCTACGGGGGATGGGAACTGTATTAGA 780  
||| |||||  
DB 1149 CTATCTTGGAAATCTCAGAATTTGGCGCTATGCTAGTGTGATGGGAGACTGATATTAG 1208  
QY 781 CCTGTAACTGACACGTGCACACAGATGTGGCAGCTCCACTGTGACATTTGCTAGCTGAA 840  
|||||  
DB 1209 CCTGTAACTGACACGTGCACACAGATGTGGCAGCTCCACTGTGACACGTGAGTGA 1268  
QY 841 GTAAAGGACAAAATTTGAGGTGTGAGTCCCATTTGTAGACAGTGTATCCCTGCT 900  
|| |||||  
DB 1269 GTAAAGGACAAAATTTGAGGTGTGAGTCCCATTTGTAGACAGTGTATCCCTGCT 1328  
QY 901 CCTCATATTTACCCCTGGCTGTCCAGAAACCTTGACATGACATTTGATAGAGTCCAT 960  
|||||  
DB 1329 CCTCATATTTACCCCTGGCTGTCCAGAAACCTTGACATGACATTTGATAGAGTCCAT 1388  
QY 961 GGTGATCTGCACTGTGTGGGTATCCCATGTTGTCACTGACTGATTCGCCACACACC 1020  
|||||  
DB 1389 GGTGACCCCTGCAGTGTGTGGGTGTCTCAGTTGTCAAAATGACTGATCCGCCACACGCC 1448  
QY 1021 TGGCTGGAAGAAATFAGAGAGGCCACCAAGACCTAGGCTCAACATCCAGTTATT 1080  
|||||  
DB 1449 TGGCTGGAAGAAATFAGAGAGGCCACCAAGACCTAGGCTCAACATCCAGTTATT 1508  
QY 1081 GGAGTCCATGTTAGACGACAGACAAAGTGGGAGCGGAGACGACCTTCATCCATTTAG 1140  
|||||  
DB 1509 GGAGTCCATGTTAGACGACAGACAAAGTGGGAGCGGAGACGACCTTCATCCATTTAG 1568  
QY 1141 GAATACAGCGTGCAGCTTGAAGAGACTTTCAGCTTCTTGCTGCGCAATGCAAGTGA 1200  
|| |||||  
DB 1569 GAGTACATGCTGATGTTGAAGAACATTTTCAGCTTCTTGCAAGCAAGATGCAAGTGA 1628  
QY 1201 AAAAAAGGTGATTTGGCCACAGATGACCCGCTTGTAAAGAGGCAAAAAACAAG 1260  
|||||  
DB 1629 AAAAAAGGTGATTTGGCCACAGATGACCCGCTTGTAAAGAGGCAAAAAACAAG 1688  
QY 1261 TACCCAGTATGATTTATAGTATACATCTCTTGGTGCAGCTGCAGTACATTAAT 1320  
|||||  
DB 1689 TACCCAGTATGATTTATAGTATACATCTCTTGGTGCAGCTGCAGTACATTAAT 1748  
QY 1321 CGATATACGAAATTCACCTTGGGGTGTGATCCGATATACATTTCTCTCCAGGCA 1380  
|||||  
DB 1749 CGATATACGAAATTCACCTTGGGGTGTGATCCGATATACATTTCTCTCCAGGCA 1808  
QY 1381 GACTTCCTAGTGTACTTTTCATGCGAGGTCTGTAGAGTTCCTATGAATCATGCA 1440  
|||||  
DB 1809 GACTTCCTAGTGTACTTTTCATGCGAGGTCTGTAGAGTTCCTATGAATCATGCA 1868  
QY 1441 GGGCTGCATCTGATGCTCTGCGAATTCCTGTTTGGATGACATCTACTATTTTGA 1500  
|| |||||  
DB 1869 ACATCTCATCTGATGCTCTGCGAATTCCTGTTTGGATGACATCTACTATTTTGG 1928  
QY 1501 GGGCCAAATGCCCAACCAATTTGCTATCTCACCAACCTTGAACTGAAGAGAA 1560  
|||||  
DB 1929 GGGCCAAATGCCCAACCAATTTGCTATCTCACCAACCTTGAACTGAAGAGAA 1988  
QY 1561 ATCCCATGGAACCTGGAGATATATTGTTGGTGGCTGGAATCATCTGGGATGGCTATCCT 1620  
|| |||||  
DB 1989 ATCCCATGGAACCTGGAGATATATTGTTGGTGGCTGGAATCATCTGGGATGGCTATCCT 2048  
QY 1621 AAAGTGTATACGAAATCTGGGAGAGCGGCTATATCCCTCTACAAAGTTTCAGAG 1680  
|||||  
DB 2049 AAAGTGTATACGAAATCTGGGAGAGCGGCTATATCCCTCTACAAAGTTTCAGAG 2108  
QY 1681 AAAGTGAAGACGTCAAGTATCCCGACATATCCCGAGGCTGACAAAGTAA 1728  
|||||  
DB 2109 AAAGTGAAGACGTCAAGTATCCCGACATATCCCGAGGCTGACAAAGTAA 2156

RESULT 9  
HSX17978  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

HS17978  
Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase  
transcript A1.  
Y17978  
Y17978.1 GI:3451266  
FUT8 gene: glycoprotein 6-alpha-L-fucosyltransferase.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 3186)  
Callieu,A., Balanzino,L., Candelier,J.J., Oriol,R. and  
Mollicone,R.  
Unpublished  
2 (bases 1 to 3186)  
Callieu,A.L.V.  
Direct Submission  
Submitted (19-AUG-1998) A.L.V. Callieu, INSERM U178, 16 av P.V.  
Courtilier, 94807 Villejuif Cedex, FRANCE  
Related entry: D89289.  
Location/Qualifiers  
1..3186  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript A1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..3162  
/gene="FUT8"  
1..389  
/gene="FUT8"  
/note="A"  
390..488  
/gene="FUT8"  
/note="C"  
489..2444  
/gene="FUT8"  
/note="D"  
717..2444  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CA76987.1"  
/db\_xref="GI:3451267"  
/translation="MRPMTGSWRWIMLIFANGTLLFYTGHLVRDNDHPDIISRELS  
KILATLERLKOONEDLRMAESLRIPBGRIDGPAIGRVRLERDLVAKKIDIEYKK  
OTRNGLGDHETLRRIENGAKELMFPQSELSKKIKNEGNLORHNADEPLDIGHNE  
RSIMTDLTYSOTDAGDMREKAEKDLTELVRITTYLQNPKDCSKAKLVQNIKGC  
GEGCOLHHVVCFLMAYGTQRTLIESONMVRATGWEVFPVSECTDWSGISTGH  
WSGEVKDRNVQVVELPIDSLHPRPYPLAVPEDLADRLVHGDPAVMVHVEHFOV  
LIRPQMLEKELEEAATKLGFEKHPVIGVHVRTRDVGTEAFAHPLEEVYHVEHFOV  
LARMQVKKRYVLTATDDPSILKEAKTYTPYEFTSDMSISGAGLHNRYSNLRGV  
LIDHIFLSQADFLVCTFSSQVCRAVEYETMOTLHPDASANFSLDITYTFCGONAHNOI  
ALYAHQPTADEIPMEPDITIGVACNHWGYSKGVNRKLGRTGLPSYKVRKLETIVK  
YPTYPEAK"

## variation

/gene="FUT8"  
/note="I"  
/frequency=".5"  
/replace="c"  
/gene="FUT8"  
/note="putative"  
3058..3063  
polyA\_signal  
polyA\_signal  
polyA\_signal

```
/gene="FUT8"
/Note="putative"
3157..3162
polyA_signal
/Note="putative"
/Note="putative"

BASE COUNT      910 a      710 c      721 g      840 t      5 others
ORIGIN

Query Match      87.4% Score 1510.4 DB 9 Length 3186;
Best Local Similarity 92.1% Pred No 0;
Matches 1592: Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Oy 1 ATGCGGCATGAGCTGTTGCTGGCGTGGATTATGCTCAATCTCTTTTGGCGGGAGCC 60
    |||||
Db 717 ATCGGCGCATGAGCTGTTGCTGGCGTGGATTATGCTCAATCTCTTTTGGCGGGAGCC 776

Oy 61 TTGCTATTTTACATAGTGTCTACTTGTACGAGATAATGACACCTGTACTCTAGC 120
    |||||
Db 777 TTGCTATTTTATATAGTGTCTACTTGTACGAGATAATGACACCTGTACTCTAGC 836

Oy 121 CGAGACTGTCCAAAGATTTTGGCAAGCTGGAGCGCTTAAACACAAATGAAGCTTG 180
    |||||
Db 837 CGAGAACTGTCCAAAGATTTTGGCAAGCTTGAACGCTTAAACACAGAAATGAAGCTTG 896

Oy 181 AGGAGATAGCTGATCTCTCCGAATACCAAGAGCCCATTTGATGAGGGCCAGCTTCA 240
    |||||
Db 897 AGGCAATGGCCGAAATCTCTCCGGAATACCAAGAGCCCATTTGATGAGGGCCAGCTTCA 956

Oy 241 GGAAGATTCGTCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTTGAAATTTAT 300
    |||||
Db 957 GGAAGATTCGTCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTTGAAATTTAT 1016

Oy 301 AAGAACAACACTTAAATGGTCCAGGGAAGGATCATGAATCCTTAAGGAGGAGATTGAA 360
    |||||
Db 1017 AAGAACAACAGCAAAATGGTCTGGGGAAGGATCATGAATCCTTAAGGAGGAGATTGAA 1076

Oy 361 AATGAGACTAAAGAGCTGTGTTTCTTCAACAAGTGAAGTTGAAGAATTTAAAGATTTA 420
    |||||
Db 1077 AATGAGACTAAAGAGCTGTGTTTCTTCAACAAGTGAAGTTGAAGAATTTAAAGATTTA 1136

Oy 421 GAAGGAATAGACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480
    |||||
Db 1137 GAAGGAATAGACTCCAAAGACATGAGATGAATTTCTTATGATTTAGGACATCATGAA 1196

Oy 481 AGGTCTATATAGAGGATCTATCTACTCTCAGTCAAAAGATGGGCGAGTGTATGGCGT 540
    |||||
Db 1197 AGGTCTATATAGAGGATCTATCTACTCTCAGTCAAAAGATGGGCGAGTGTATGGCGG 1256

Oy 541 GAAAAGGAGGCCAAAGATCTGACAGAGCTGTCCAGCGGAGATTAACATATCTTCAGAA 600
    |||||
Db 1257 GAAAAGGAGGCCAAAGATCTGACAGAGCTGTTCAGCGGAGATTAACATATCTTCAGAA 1316

Oy 601 CCCAAGGATCGCAAGAGCCAGAGCTGATGTAATATCAACAAGGCTGTGGCTAT 660
    |||||
Db 1317 CCCAAGGATCGCAAGAGCCAGAGCTGTGTAATATCAACAAGGCTGTGGCTAT 1376

Oy 661 GGCTGTACAGCTCATATAGTACTGATCTTATGATTTGACATATGGACCCAGCAACA 720
    |||||
Db 1377 GGCTGTACAGCTCATATAGTACTGATCTTATGATTTGACATATGGACCCAGCAACA 1436

Oy 721 CTCGCTTGGGAATCTCACAATTTGGCGCTACGCTACTGCGGGGATGGGAAACTGTGTTAGA 780
    |||||
Db 1437 CTCGCTTGGGAATCTCAGAATTTGGCGCTATGCTACTGTGGATGGGAACTGTATTTAGG 1496

Oy 781 CCTGTAAGTAGACGCTGACACAGACATCTGGGAGCTCCACTGGACATTTGTTAGGTGAA 840
    |||||
Db 1497 CCTGTAAGTAGACGCTGACACAGACATCTGGGAGCTCCACTGGGAGCTGTGTTAGGTGAA 1556

Oy 841 GTAAGGACAAAAATGTTACAGGTGTTGAGAGTCCCATTTGTACAGCTGTTTCCCTCGT 900
    |||||
Db 1557 GTAAGGACAAAAATGTTCAAGTGTGTGAGAGTCCCATTTGTACAGCTGTTTCCCTCGT 1616

Oy 901 CCTCATATTTTACCCCTGTGCTGCCAGAAAGACCTTGAAGATGAGCTTGTAGAGTGCAT 960
```

```
|||||
Db 1617 CTTCCATATTTTACCCCTGTGCTGCCAGAAAGACCTGCGAGATGACTGTAGAGTGCAT 1676

Oy 961 GGTGATCCGCAAGTGTGGGTGATATCCAGTTTGTGTAAGTACTGTTGGCCACACAACC 1020
    |||||
Db 1677 GGTGATCCGCAAGTGTGGGTGATATCCAGTTTGTGTAAGTACTGTTGGCCACAGCCCT 1736

Oy 1021 TGGCTGGAAGAAATAGAGAGAGCCACCAAGAGACTGAGCTTCAACATCCAGTTAT 1080
    |||||
Db 1737 TGGCTGGAAGAAATAGAGAGAGCCACCAAGAGACTGAGCTTCAACATCCAGTTAT 1796

Oy 1081 GGAATCCATGTTAGACGACACAGCAAGTGGAGCGGAGACGCTTCCATCCATTGAG 1140
    |||||
Db 1797 GGAATCCATGTTAGACGACACAGCAAGTGGAGCGGAGACGCTTCCATCCATTGAG 1856

Oy 1141 GAATACAGCTGACGCTTGAAGAGACTTACCTCTTGGTCCGAGATTCGAAGTGGAT 1200
    |||||
Db 1857 GAATACAGCTGACGCTTGAAGAGACTTACCTCTTGGTCCGAGATTCGAAGTGGAT 1916

Oy 1201 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTGTTAAAGAGCCAAACAAAG 1260
    |||||
Db 1917 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTGTTAAAGAGCCAAACAAAG 1976

Oy 1261 TACCCAGTTATGATTTATTTAGTATTAACCTATCTTGTGTGACCTGACTACATTAAT 1320
    |||||
Db 1977 TACCCAGTTATGATTTATTTAGTATTAACCTATCTTGTGTGACCTGACTACATTAAT 2036

Oy 1321 CGATATACGAAATTCACCTTGGGGTGTATCCGAGATTAACCTTCTCCCGAGGCA 1380
    |||||
Db 2037 CGATATACGAAATTCACCTTGGGGTGTATCCGAGATTAACCTTCTCCCGAGGCA 2096

Oy 1381 GACTTCTAGTGTACTTTTTCATGCGAGGCTGTAGAGTTGCTTATGAAATCATGCAA 1440
    |||||
Db 2097 GACTTCTAGTGTACTTTTTCATGCGAGGCTGTAGAGTTGCTTATGAAATCATGCAA 2156

Oy 1441 GCGCTCATCTGATGATCCCTCTGCGAACTTCCTGTTTGGATGACATCTACTATTTTGA 1500
    |||||
Db 2157 ACACATCATCTGATGATCCCTCTGCGAACTTCCTGTTTGGATGACATCTACTATTTTGA 2216

Oy 1501 GGCCCAATGCCACACAACCAAAATTTGCTATTAATCCACCAACCTGAGCTGAAGAGAA 1560
    |||||
Db 2217 GGCCCAATGCCACACAACCAAAATTTGCTATTAATCCACCAACCTGAGCTGAAGAGAA 2276

Oy 1561 ATCCCATGGAACCTGAGATTTATTTGGTGTGCTGTAAGAACTCTGGATGCTATCTCT 1620
    |||||
Db 2277 ATCCCATGGAACCTGAGATTTATTTGGTGTGCTGTAAGAACTCTGGATGCTATCTCT 2336

Oy 1621 AAAGTGTTAACGAAACTGGGAGAGAGGCTTATTCCTCTTCAAAAGTTCCAGAG 1680
    |||||
Db 2337 AAAGTGTTCACAGAAATTTGGGAGAGAGGCTTATTCCTCTTCAAAAGTTCCAGAG 2396

Oy 1681 AAGATGAACAGTCAAGTACCCACATATCCGAGAGCTGACAGTA 1728
    |||||
Db 2397 AAGATGAACAGTCAAGTACCCACATATCTGAGGCTGAGAAATTA 2444

RESULT 10
HSY17979      3280 bp      mRNA      linear      PRI 06-NOV-1998
LOCUS      HSY17979
DEFINITION      Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase
transcript A2.
ACCESSION      Y17979.1 GI:3451268
VERSION      Y17979.1
KEYWORDS      FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 3280)
AUTHORS      Calileu A., Balanzino L., Candeller J.J., Oriol R. and
Mollicone R.
TITLE      Differential splice variants of human FUT8 embryonic cDNA
JOURNAL      Unpublished
```

REFERENCE 2 (bases 1 to 3280)  
AUTHORS Caillieu,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Caillieu, INSERM U178, 16 av P.V.  
Coururier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source 1..3280  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript A2"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..3162  
/gene="FUT8"  
1..389  
/gene="FUT8"  
/note="A"  
390..488  
/gene="FUT8"  
/note="C"  
489..2444  
/gene="FUT8"  
/note="D"  
717..2444  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele I"  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76988.1"  
/db\_xref="GI:3451269"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPTGSWRIMLILFAGTLLFYIGHLVDRNDHPDHSRRLS  
KILAKLERKQONEDLRMAESLRIPEDPIDGPAIGRVRLLEPOLYKAKKEIDENK  
OTRNLGDHEILRRRIENGAKELMFLOSEKLLKLEGNLORADFLDLCHHE  
RSINTDLVLSOTDAGDKREKBAKDTELVORITTYLONPKDCSAKLVGNINSG  
GYGQDLHVYCFMAYGTQRTLLESOMRYATGMEVFRPVSCTDRSGISG  
MSGEVKKVQVLEPLVDSLHPRPYLPYLAPEDLADLVHGDPAVMWYSEVYK  
LIRPQMLEKEIEEATKLGFKHPVIGVHVRDVKGTFAAHPHIEEYVHEHRL  
LARMOVDKKRYLALTDPSILKEAKTKYPNEYISDSAGIHNYTENSLRGV  
LIDHIFLSADPLVCTFSSQYCRVAYELMOTLHPDASNFHSLDIYYGGOMAHQI  
ALYAHQPTADEIPMEPGDIGVAGNMDGYSKGNKRLGTLPSYKREKIEYK  
LYYPEAEK"  
2506  
variation  
/gene="FUT8"  
/note="I1"  
/frequency=".5"  
/replace="C"  
3046..3051  
/gene="FUT8"  
/note="putative"  
3058..3063  
/gene="FUT8"  
/note="putative"  
3157..3162  
/gene="FUT8"  
/note="putative"  
BASE COUNT 934 a 718 c 743 g 880 t 5 others  
ORIGIN  
Query Match 87.4%; Score 1510.4; DB 9; Length 3280;  
Best local Similarity 92.1%; Pred. No. 0;  
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
QY 1 ATGCGGCATGAGCTGCTGCGGCTGATGATGCTCTTTTGGCTGGGGACC 60  
DB 717 ATGCGGCATGAGCTGCTGCGGCTGATGATGCTCTTTTGGCTGGGGACC 776  
QY 61 TTGCTATTTTACATAGGTGCTACTGTGACGATATGACCACCTGATCACTAGC 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 777 TTGCTGTTTATATAGGTGCTACTGTGTAACGATATGACCATCTGATCACTAGC 836  
QY 121 CGAGAACTGTCACAAATTTTGGCAAGCTGGAAACCTTAAACAAACAAATGAAGCTTG 180  
DB 837 CGAGAACTGTCACAAATTTTGGCAAGCTGGAAACCTTAAACAAACAAATGAAGCTTG 896  
QY 181 AGGAGAACTGCTGATCTCTCCGATACAGAAAGGCCCCATGATGAGGGGCCCATCTCA 240  
DB 897 AGGCAATGGCCGAATCTCTCCGATACAGAAAGGCCCCATGATGAGGGGCCCATCTCA 956  
QY 241 GGAAGAGTTCGTGCTTTAGAGAGCAATTTATGAAAGGCCCAAGAACAGATTGAAATTTAT 300  
DB 957 GGAAGAGTTCGTGCTTTAGAGAGCAATTTATGAAAGGCCCAAGAACAGATTGAAATTTAT 1016  
QY 301 AGAACAACAACTAAATATGCTCCAGGAAAGAGATCATGAATTCCTAAGAGAGAGATTGAA 360  
DB 1017 AAGAAACAGACCCAGAAATGCTCGGGGAAAGATCATGAATTCCTGAGAGAGATTGAA 1076  
QY 361 AATGAGCTTAAGAGCTCGGTTTCTTACAAAGTGAATTTGAAAGAAATTTAAAGAAATTTA 420  
DB 1077 AATGAGCTTAAGAGCTCGGTTTCTTACAAAGTGAATTTGAAAGAAATTTAAAGAACTTA 1136  
QY 421 GAAGAAATGAACCTCCAAAGACATGCAGATGAATTTCTATCAGATTTGGACATCATGAA 480  
DB 1137 GAAGAAATGAACCTCCAAAGACATGCAGATGAATTTCTTTTGTGATTTAGACATCATGAA 1196  
QY 481 AGGTCTATATGACGATCTATCTACCTCAGTCAACAGATGGGAGCTGATTTGGCGT 540  
DB 1197 AGGTCTATATGACGATCTATCTACCTCAGTCAACAGATGGGAGCTGATTTGGCGG 1256  
QY 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCACAGGGAGATTAACATATCTTCAGAA 600  
DB 1257 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCACAGGGAGATTAACATATCTTCAGAA 1316  
QY 601 CCCAAGAGCTGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 660  
DB 1317 CCCAAGAGCTGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1376  
QY 661 GCGTCTACGCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
DB 1377 GCGTCTACGCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436  
QY 721 CTGCGCTGGAATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
DB 1437 CTGATCTGGAATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1496  
QY 781 CCTGTAGTGAAGAGCTGACAGACAGATGCGACGCTCCACTGACATTTGGTCAGGTGAA 840  
DB 1497 CCTGTAGTGAAGAGCTGACAGACAGATGCGACGCTCCACTGACATTTGGTCAGGTGAA 1556  
QY 841 GTAAAGACAAAATGTTTCAGGTGTTGAGTCCCATTTGTACAGATGTTCAATCCTCGT 900  
DB 1557 GTAAAGACAAAATGTTTCAGGTGTTGAGTCCCATTTGTACAGATGTTCAATCCTCGT 1616  
QY 901 CCTGCATATTTACCCCGGCTGTCACAAAGACCTTGCAATGATGATGATGATGATGATGATGAT 960  
DB 1617 CCTGCATATTTACCCCGGCTGTCACAAAGACCTTGCAATGATGATGATGATGATGATGATGAT 1676  
QY 961 GGTGATCTGCTAGTGTGTTGGGTATCCAGTTTGTCAAGTACTTGTATTCGCCCAACACC 1020  
DB 1677 GGTGATCTGCTAGTGTGTTGGGTATCCAGTTTGTCAAGTACTTGTATTCGCCCAACACC 1136  
QY 1021 TTGCTGGAAGAAAGAAATAGAGAGGCCACCAAGAGCTAGGTTCAACATTCAGTTATT 1080  
DB 1737 TTGCTGGAAGAAAGAAATAGAGAGGCCACCAAGAGCTAGGTTCAACATTCAGTTATT 1796  
QY 1081 GGAATCCATGTTAGAGCAGACAGCAAAAGTGGAGGGAACACAGCTTCCATCCATTTAG 1140  
DB 1797 GGAATCCATGTTAGAGCAGACAGCAAAAGTGGAGGGAACACAGCTTCCATCCATTTAG 1856  
QY 1141 GAATACACGGTGCACGTTGAAGAAGACTTTCACCTTCTGCGAGAAATGAGTGAAT 1200  
DB 1857 GAATACACGGTGCACGTTGAAGAAGACTTTCACCTTCTGCGAGAAATGAGTGAAT 1916





OY	1261	TACCCAGTTATGAAATTTATAGTATGAATACCTATCTCTGGTACGCTGACATACATAT	1320
Db	1458	TACCCCAATTTATGAATTTATATAGTATACCTATTTCTCGTACGCTGATGCACAT	1517
OY	1321	CGATATACGAAAAATTCCTACTTCGGGGTGTGATCCGTGATATACACTTTCTCTCCAGGCA	1380
Db	1518	CGATACACGAAAAATTCCTACTTCGTGGAGTGAATCCGTGATATACATTTTCTCTCAGGCA	1577
OY	1381	GACTTCCTTAGTGTGTACTTTTTCATCGCAGGTCTGTAGAGTTGCTTATGAAATCATGCA	1440
Db	1578	GACTTCCTTAGTGTGTACTTTTTCATCCAGGCTCTCGAGTGTCTTATGAAATTTATGCA	1637
OY	1441	GGCGTCATCTCGATGAGCCCTCGCAGAACTCCGTCCTGTTGGATGACATCTACATATTTGGA	1500
Db	1638	ACACTACATCTCGATGAGCCCTCGCAGAACTCCGTCCTGTTGGATGACATCTACATATTTGGG	1697
OY	1501	GGCCCAAAATGCCCAACCAAAATTCGCATTTATCTCCACCAACTCTGAACTGAAGAGAA	1560
Db	1698	GGCCAGATATGCCCAACCAAAATTCGCATTTATCTCCACCAACCCGGAACCTCAATGAA	1757
OY	1561	ATCCCCATGGAAACCTGGAGATATTTATGCTGTGGCTGGGAAATCACTGGAGTGGCATCTT	1620
Db	1758	ATTCCTCATGGAACCTGGAGATATTCATTTGTGTGGCTGGGAAATCATTTGGATGGCATTTCT	1817
OY	1621	AAAGGTGTTTAAACAGAAAACTGGGAAAGGAGCGGCTATATCTCTCTACAAAGTTTGAGAG	1680
Db	1818	AAAGGTGTTTAAACAGAAATTTGGGAAGGAGCGGCTATATCTCTCTACAAAGTTTGAGAG	1877
OY	1681	AAAGATGAAACAGTCAAGTACCTCCCATATATCCGAGGCTGACAAAGTAA	1728
Db	1878	AAAGATGAAACAGTCAAGTACCTCCCATATATCTCTGAGGCTGACAAAGTAA	1925

FEATURES	REMARK	COMMENT
LOCUS	BC010666	
DEFINITION	BC010666 . 2976 bp mRNA	linear
ACCESSION	mus musculus, fucoseyltransferase 8, clone MGC:11418	IMAGE:3594582,
VERSION	BC010666	
KEYWORDS	mRNA, complete cds.	
SOURCE	BC010666.1 GI:14715012	
ORGANISM	MGC.	
	house mouse.	
	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 2976)	
TITLE	Strausberg, R.	
JOURNAL	Direct Submission	
	Submitted (10-JUL-2001) National Institutes of Health, Mammalian	
	Gene Collection (MGC), Cancer Genomics Office, National Cancer	
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,	
	USA	
	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>	
	Contact: MGC help desk	
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>	
	Tissue Procurement: Jeffrey Green M.D.	
	cDNA Library Preparation: Life Technologies, Inc.	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	
	DNA Sequencing by: Baylor College of Medicine Human Genome	
	Sequencing Center	
	Center code: BCM-HGSC	
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>	
	Contact: <a href="mailto:amgdbcm.tmc.edu">amgdbcm.tmc.edu</a>	
	Gunaratne, P.H., Garcia, A.M., Lu, X., HuiYK, S.W., Hale, S.M.,	
	Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,	
	Richards, S., Gibbs, R.A.	
	Clone distribution: MGC clone distribution information can be found	
	through the I.M.A.G.E. Consortium/LNL at: <a href="http://Image.lnl.gov">http://Image.lnl.gov</a>	
	Series: IRAX Plate: 16 Row: 1 Column: 5	
	This clone was selected for full length sequencing because it	
	passed the following selection criteria: matched mRNA gi.:8593370.	
	Location/Qualifiers	

[illegible]

OY	601	CCCAAGSACGTGCAGCAAAAGCCAAAGCAGCTAGTGTGAATATATCAAAAGSGCTGGGCTAT	660
Dp	1184	CCTAAGSACGTGCAGCAAAAGCCAAAGCAGCTAGTGTGAATATATCAAAAGSGCTGGGCTAT	1243
OY	661	GGCGTGCAGCTCCATCATGTAGTGTACTCCTTTATGATTGCATATGACACCCAGCGAACA	720
Dp	1244	GGTTTGCAGCTCCATCATCGTGGTCTACTGCTTTTCATGATTGCTTATGGCACCCAGGAACA	1303
OY	721	CTCGCCTTGGAAATCTCACAAATTGGCCCTACGCTACTGGGGGATGGGAAACTGTGTTTAGA	780
Dp	1304	CTCATCTTGGAAATCTCAGAAATTGGCCCTATGCTACTGGGTGATGGGAACATGTGTTTAGA	1363
OY	781	CCCTAATAGTAGAGCTGTGCACAGACAGATATGTGCAGCTCCACATGAGACATTTGGCAGGTAA	840
Dp	1364	CCTTAAAGTAGAGACATGTGTACAGACAGATATGTGCAGCTCCACATGAGACACTGGTCAAGTAA	1423
OY	841	GTAAGAACAATAATGTTCAAGTGTGGTGTGAAGCTCCCAATTTGAGACAGTTCATCTCTGT	900
Dp	1424	GTAATATGACAAAAAATTCATCAAGTGTGTGCAGCTCCCATTTGTAGACAGCTCCATCTCTCG	1483
OY	901	CCCTCATATTACCCTCGGCTGTGCCAGAAAGCCTTGCAGATGCAAGCTGTACGAGTCCAT	960
Dp	1484	CCCTCCTTACTTACCATCGCTGTTCACAGAAAGCCTTGACAGACCGCACTCCTTAAGGTTCAT	1543
OY	961	GGTGATCCTGCAGTGTGTGGTGGTATCCCAAGTTCATGACTTGTATTGGCCCAACACC	1020
Dp	1544	GGTGACCTCGAGTGTGTGGTGGTGTCCCAAGTTCATGACTTGTATTGGTCCCAACACT	1603
OY	1021	TGCGTGGAAAAAGAAATATAGAAGAGGCCACCAAGAACTATAGCTTAAACATCCAGTTATT	1080
Dp	1604	TGCGTGGAAAAAGAAATATAGAAGAGGCCACCAAGAACTATAGCTTAAACATCCAGTTATT	1663
OY	1081	GGAGTCCATGTTAGACGCACAGACAAAGTGGGAGGGAAGCAGCTTCATCCATCCATTGAG	1140
Dp	1664	GGAGTCCATGTCAGAGGSCACAGACAAAGTGGGAGGGAAGCAGCTTCACACCATCGAG	1723
OY	1141	GAAATACACGGTGCACGTTGAAGAAGCTTTCAAGCTTTCTTGCTGCAGAAATGCAGTGGAT	1200
Dp	1724	GAGTACATGTGACACGTTGAAGAAGCAATTTTCAGCTTCTCGCACGAGATCAGAGTGGAT	1783
OY	1201	AAAAAAGGGGTATTGGGCAACAGATAGACCGGCTTTGTTAAAGAGGCAAAAAAAG	1260
Dp	1784	AAAAAAGAGATATATCTGCGCTACTGATGATCCTACTTTGTTAAAGAGGCAAAAGCAAG	1843
OY	1261	TACCCAGTTATGAATTTATTTAGTGAATACCTATCTCTTGGTCAGGTGACATATAT	1320
Dp	1844	TATCTCAATTTATGAATTTATTTAGTGAATACCTATCTCTTGGTCAAGCTACACAT	1903
OY	1321	CGAATATACAAAAATTCATCTTCGGGGTGTGATTCGTGGATATACACTTCTCTCCAGGCA	1380
Dp	1904	CGGTACACAGAAAAATTCATCTTCGGGGTGTGATTCGTGGATATACACTTCTCTCACAGGCT	1963
OY	1381	GACTTCCTAGTGTACTTTTTCATGCGAGAGTCTGTAGAAGTGTCTTAAATCATGCA	1440
Dp	1964	GACTTCTAGTGTACTTTTTCATGCCAGGTCTGTGGGGTGTCTTAAATCATGCA	2023
OY	1441	CGCGTCATCTGTATGCTCTGTGGAACTTCGTTCTTTGGATGACATCTACTATTTTGA	1500
Dp	2024	ACCTGTGATCCTGTATGCTCTGTGGAACTTCATCTTTTGGATGACATCTACTATTTTGA	2083
OY	1501	GGCCCAATGGCCCAACCAATTTGGCATTTATCTCCACACACCTGTGAAAGGAGAA	1560
Dp	2084	GGCCCAATGGCCCAACCAATTCAGATTGCTGTTTATCTCCACAACTGTGAAAGGAGAA	2143
OY	1561	ATCCCATGGAACCTGTGAGATATTTATTTGTTGTGGCTGTGAAATCACTGGGATGGCTATCT	1620
Dp	2144	ATTCCATATGGAACCTGTGAGATATCATTTGTTGTGGCTGTGAAACCATTTGGATGATTATCT	2203
OY	1621	AAAGGTGTAAACAGAAACCTGTGGGAAGAGGGGCTATATCCTCTCAAAAGTTCGAGAG	1680
Dp	2204	AAAGGTGTAAACAGAAACCTTGTGAAACAGAGCTGTATATCTCTCTCAAAAGTTCGAGAG	2263

RESULT	13
AB025198	2176 bp mRNA linear ROD 05-AUG-2000
LOCUS	
DEFINITION	Mus musculus mRNA for alpha-1,6-fucosyltransferase, complete cds.
ACCESSION	AB025198
VERSION	AB025198.1 GI:4586553
KEYWORDS	alpha-1,6-fucosyltransferase.
SOURCE	Mus musculus adult male cerebrum cDNA to mRNA.
ORGANISM	Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
AUTHORS	1 (sites)
TITLE	Hayashi, H., Yoneda, A., Asada, M., Ikeita, M. and Imamura, T.
JOURNAL	Molecular cloning of mouse alpha-1,6-fucosyltransferase and expression of its mRNA in the developing cerebrum
MEDLINE	DNA Seq. 11 (1-2), 91-96 (2000)
REFERENCE	20358720
AUTHORS	2 (bases 1 to 2176)
TITLE	Hayashi, H., Yoneda, A. and Imamura, T.
JOURNAL	Direct Submission
FEATURES	Submitted (19-MAR-1999) Toru Imamura, National Institute of Bioscience and Human Technology, Biosignaling Department; 1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan (E-mail: Imamura@nih.go.jp, Tel:81-298-54-6072, Fax:81-298-54-6149)
SOURCE	Location/Qualifiers
	1..2176
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/sex="male"
	/tissue_type="cerebrum"
	/dev_stage="adult"
	205..1932
	/codon_start=-1
	/product="alpha-1,6-fucosyltransferase"
	/protein_id="PAA76392.1"
	/db_xref="GI:4586554"
	/translation="MRAMTSMRIMLIFAMGTLIFYIGHLVRDNDIPDHSTRELS KILAKLERLQONDEIRMAAESLRIPGSPIDOCATGCRVAVLEOLYAKAEOLENYKK QARNGLEKDEHLERRLENGKKELELFPLOSELKKLKLKEGNELOHADDELIDLIGNE RSITMDLYVLSDTDAGDMREKKEKNDLYQRTYLVLPMPKDCSKARKLVGNKGC GYGCOLHHVYLCFMIAIGTORLTILLESQNNRYAAGMETVFRVSEETCDRSGLSGH WSGVNDKNIOVLELPRLVDSLFHPRPYPLPAVEDLADRLRVHGDPVAMVSOFYK LIRPQPLKEKEIEEATKKLGFKHPVIGVHVRIDKVGCTEAAPFPIEEYVHVEHQFL LARMQYDKRKRYLATDPTLILKANTKYSNYERISDNTSISMSAGLHNRYTENSRLV ILDLHLSQADFLVCTFSVQYCAVAVETMOTLHVDASANTHSLLDYLFGQNNHNDI AVVHFKPRTEELIMPEGDIIIGVGNHMDYSKQINRKLKGTGLYPSYKVRKEIETVK YPRYPEAK"
BASE COUNT	669 a 474 c 494 g 539 t
ORIGIN	
Query Match	84.8% Score 1464.6; DB 10; Length 2176;
Best Local Similarity	90.5%; Pred. No. 0;
Matches 1563; Conservative	0; Mismatches 164; Indels 0; Gaps 0;
OY	1 ATGCGGCATGAGCTGGTTGGTGGCGGTGAGATTAGCTCATCTTTTGGCTGGGCGACC 60
Db	205 ATGCGGCATGAGCTGGTTGGTGGCGGTGAGATTAGCTCATCTTTTGGCTGGGCGACC 264
OY	61 TTGCAATTTTACATGGAGTGCATCTTGGTAGAGATAATGACCACTGTGACACTCTAGC 120
Db	265 TTGTTATTTTATATGAGTGTGATTTGGTTGGAGATAATGACCACTGTGACACTCTAGC 324
OY	121 CGAGAACTGTCCAAAGATTGTTGGCAAAGCTGGAACGCTTAAACAACAATAATGAAGACTTG 180
Db	325 AGAGAACTGTCCAAAGATTCTTGCAAAAGCTTGAAGCGCTTAAACAACAATAATGAAGACTTG 384
OY	181 AGGAGATAGCTGATCTCTCCGAATACCAAGAGGCCCATTTGATCAGGGGCGAAGCTTCA 240

```

1 Db 385 AGCGAATGGCTAGTCTCTCCGAATACCAGAGGCCCATTTGACCAGGGACAGCTACA 444
Oy 241 GGAAGAGTTCGTCCTTTAGAGCAATTTATGAAGGCCCAAGAACAGATTTGAAATTTAT 300
Db 445 GGAAGAGTCCGCTGTTTACAGAGACAGCTTGTAAAGGCCCAAGAACAGATTTGAAATTTAT 504
Oy 301 AAGAAACAACATTAATAATGTCAGGAGAGATCATGAATCTTAAGAGGAGATTGAA 360
Db 509 AAGAAACAACATTAATAATGTCAGGAGAGATCATGAATCTTAAGAGGAGATTGAA 564
Oy 361 AATGAGCTAAAGAGCTGCTGTTTTTCTACAAAGTACTGAAGAATTTAAAGATTTTA 420
Db 565 AATGAGCTAAAGAGCTGCTGTTTTTCTACAAAGTACTGAAGAATTTAAAGATTTTA 624
Oy 421 GAAGAAATGAAGCTCCAAAGACATGAGATTAATTTCTATGATTTGGACATCATGAA 480
Db 625 GAAGAAATGAAGCTCCAAAGACATGAGATTAATTTCTATGATTTGGACATCATGAA 684
Oy 481 AGGTCTATATGAGCGATCTATATCTACCTCAACAGATGGGAGAGTATGGCGT 540
Db 685 AGGTCTATATGAGCGATCTATATCTACCTCAACAGATGGGAGAGTATGGCGT 744
Oy 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGCTCAGCGGAGAAATACATATCTTCAGAA 600
Db 745 GAAAGAGAGCCAAAGATCTGACAGAGCTGCTCAGCGGAGAAATACATATCTTCAGAA 804
Oy 601 CCCAAGAGCTGAGCAAGGCCAAGAAAGTATGTTGAATATACAAAGGCTGGCTAT 660
Db 805 CTTAAGAGCTGAGCAAGGCCAAGAAAGTATGTTGAATATACAAAGGCTGGCTAT 864
Oy 661 GCGTCTCAGCTCCATCATATGATGATCTATGATTTGATTCATATGAGCACCAGCAACA 720
Db 865 GCGTCTCAGCTCCATCATATGATGATCTATGATTTGATTCATATGAGCACCAGCAACA 924
Oy 721 CTCGCCCTTGAATCTCAGAAATGGCGCTACGCTACTGGGAGATGGGAAATCTGTTTGA 780
Db 925 CTCGCCCTTGAATCTCAGAAATGGCGCTACGCTACTGGGAGATGGGAAATCTGTTTGA 984
Oy 781 CCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 840
Db 985 CCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1044
Oy 841 GTTAAAGCAAAATATGTCAGTGTGTTGAGTCCCATTTGTAGACAGTGTTCCTCGT 900
Db 1045 GTTAAAGCAAAATATGTCAGTGTGTTGAGTCCCATTTGTAGACAGTGTTCCTCGT 1104
Oy 901 CCTCATATTTTACCCCTGCTGCCAGAGACCTTGGACATCTTACAGATTCAT 960
Db 1105 CCTCATATTTTACCCCTGCTGCCAGAGACCTTGGACATCTTACAGATTCAT 1164
Oy 961 GGTGATCCGTCAGTGTGTTGAGTGTGTTGAGTGTGTTGAGTGTGTTGAGTGTGTTGAG 1020
Db 1165 GGTGATCCGTCAGTGTGTTGAGTGTGTTGAGTGTGTTGAGTGTGTTGAGTGTGTTGAG 1224
Oy 1021 TGGCTGAAAGAGAAATAGAGAGGCGACCAAGAGCTAGGCTTCAACATCCAGTTAT 1080
Db 1225 TGGCTGAAAGAGAAATAGAGAGGCGACCAAGAGCTAGGCTTCAACATCCAGTTAT 1284
Oy 1081 GGAATTCATGTTAGAGCAGACAGAAAGTGGAGGCGAAGACCTTCATCCATTTGAG 1140
Db 1285 GGAATTCATGTTAGAGCAGACAGAAAGTGGAGGCGAAGACCTTCATCCATTTGAG 1344
Oy 1141 GAATACAGGCTGACGTTGAAGAGACCTTCATCCATCCATCCATCCATCCATCCATCCAT 1200
Db 1345 GAGTCAATGTTAGAGCAGTGTGAACATTTTCAAGCTTCTGACAGCAAGTGAAGTGAAT 1404
Oy 1201 AAAAAAGGCTGATTTGGCCACAGATGACCTGCTTTTAAAGAGCAAAAAACAAG 1260
Db 1405 AAAAAAGGCTGATTTGGCCACAGATGACCTGCTTTTAAAGAGCAAAAAACAAG 1464
Oy 1261 TACCCAGTATGAAATTTATGATGAATGATATCTCTTGGTCAAGCTGACATTAAT 1320
Db 1465 TACCCAGTATGAAATTTATGATGAATGATATCTCTTGGTCAAGCTGACATTAAT 1524

```

```

Oy 1321 CGATATACGAAAAATTCACCTTGGGGTGTGATCTCGATATACACTTTCCTCCAGCA 1380
Db 1525 CGGTACACAGAAAAATTCACCTTGGGGTGTGATCTCGATATACACTTTCCTCCAGCA 1584
Oy 1381 GACTTCCTAGTGTGATCTTTCATGCGAGCTGTGAGATGCTTATGAATCATGCA 1440
Db 1585 GACTTCCTAGTGTGATCTTTCATGCGAGCTGTGAGATGCTTATGAATCATGCA 1644
Oy 1441 GCGCTGATCTGATGCTCTGCGAAGCTTCCGTTCTTTGGATGACATCTATTTTGA 1500
Db 1645 ACCCTGATCTGATGCTCTGCGAAGCTTCCGTTCTTTGGATGACATCTATTTTGA 1704
Oy 1501 GGGCCAAATGCCCAACAGCAAAATGCGATTTATCCACCAACCTCGAAGCAAGGAA 1560
Db 1705 GGGCCAAATGCCCAACAGCAAAATGCGATTTATCCACCAACCTCGAAGCAAGGAA 1764
Oy 1561 ATCCCATGGAACCTGAGATATTTATGTTGGTGTGAGTGAATTCATGAGATGCTTCT 1620
Db 1765 ATTCATGGAACCTGAGATATTTATGTTGGTGTGAGTGAATTCATGAGATGCTTCT 1824
Oy 1621 AAGGTGTTACAGAAAACTGGGAGAGCGGCTATATCTCTTACAAAGTTGAGAG 1680
Db 1825 AAGGTGTTACAGAAAACTGGGAGAGCGGCTATATCTCTTACAAAGTTGAGAG 1884
Oy 1681 AAGATGAAGACGTCATATACCCCATATCCCGAGGCTGACAGTA 1727
Db 1885 AAGATGAAGACGTCATATATCCCATATCTGAAAGCTGAAATA 1931

RESULT 14
BC025385
LOCUS BC025385 2682 bp mRNA linear PRI 07-AUG-2002
DEFINITION Homo sapiens, clone MGC:26465 IMAGE:4838097, mRNA, complete cds.
ACCESSION BC025385
VERSION BC025385.1 GI:22135669
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2682)
AUTHORS Strausberg, R.
TITLE Direct Submision
JOURNAL Submitted (05-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA.
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL bc: http://image.llnl.gov
Series: IRM Plate: 34 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758407.
FEATURES
source
1..2682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:26465 IMAGE:4838097"
/tissue_type="testis"
/clone_1ib="NIH_MGC_97"

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 : Search time 31 Seconds  
(without alignments)  
3208.777 Million cell updates/sec

Title: US-09-839-136-2  
Perfect score: 3096  
Sequence: 1 MRPWGSMRWIMILIFAMGT.....YKVKRIETVKYPTPEADK 575

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeophages:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3088	99.7	575	6	P79282 sus scrofa
2	2949	95.3	575	4	O00235 homo sapien
3	2945	95.1	575	6	O9N0M2 bos taurus
4	2903	93.8	575	11	O921U1 mus musculu
5	2892	93.4	575	11	O9WTS2 mus musculu
6	1431.5	46.2	619	5	O9YVY5 drosophila
7	1420	45.9	276	4	O9P2U6 homo sapien
8	1109	35.8	246	4	O9BYC6 homo sapien
9	948	30.6	818	5	O16882 caenorhabd1
10	529	17.1	105	4	O9BYC5 homo sapien
11	261	8.4	50	4	O9P2U5 homo sapien
12	140	4.5	1025	10	O9SAAS arabidopsis
13	123.5	4.0	1057	16	O9N0R5 staphylococ
14	123	4.0	1871	10	O9SRH5 arabidopsis
15	122	3.9	1200	11	O921B9 mus musculu
16	121.5	3.9	1154	16	O9ZK97 hellobacte

17	119	3.8	1455	5	O9V587 drosophila
18	117.5	3.8	726	4	O15624 homo sapien
19	117.5	3.8	2363	4	O99968 homo sapien
20	117	3.8	1958	5	O96062 dugesia jap
21	116	3.7	544	5	O61196 caenorhabd1
22	116	3.7	1146	11	O9WVE1 ratius norv
23	116	3.7	1217	11	O9WVE9 ratius norv
24	115	3.7	584	17	O96Y06 sulfolobus
25	113.5	3.7	1097	5	O9VIE7 drosophila
26	113	3.6	1391	11	O922J3 mus musculu
27	112.5	3.6	647	5	O27031 theileria p
28	112	3.6	978	16	O9E9T3 staphylococ
29	112	3.6	1527	11	O9E867 ratius norv
30	111	3.6	398	10	O49583 arabidopsis
31	111	3.6	570	17	O97U11 sulfolobus
32	111	3.6	1011	5	O61639 drosophila
33	111	3.6	1094	5	O61618 drosophila
34	111	3.6	1549	5	O20060 caenorhabd1
35	111	3.6	1680	5	O9U8G1 plasmodium
36	111	3.6	1940	5	O02456 schistosoma
37	110.5	3.6	794	4	O96CM4 homo sapien
38	110.5	3.6	909	1	O57703 acidianus a
39	109.5	3.5	975	13	O98TQ5 notochenia
40	109.5	3.5	1179	17	O59462 pyrococcus
41	109.5	3.5	1397	10	O9LP90 arabidopsis
42	109	3.5	217	4	O961T4 homo sapien
43	109	3.5	417	17	O9V104 pyrococcus
44	109	3.5	822	17	O27154 methanother
45	109	3.5	5362	2	O93155 bacillus su

#### ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	575 AA.
ID P79282			
AC P79282:			
DT 01-JUN-1998 (TREMBLrel. 06, Created)			
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE ALPHA-(1,6)-FUCOSYLTRANSFERASE (EC 2.4.1.68) (GDP-FUCOSE--GLYCOPROTEIN DE FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 9).			
OS Sus scrofa (pig).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX NCBI_TaxID=9623;			
RM [1]			
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC TISSUE=BRAIN:			
RX MEDLINE=97066976; PubMed=6910378;			
RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X.,			
RA Teshima T., Fujii S., Shiba T., Taniguchi N.;			
RT "Purification and cDNA cloning of porcine brain GDP-L-Fuc:N-acetyl-			
RT beta-D-glucosaminide alpha1->6fucosyltransferase.";			
RL J. Biol. Chem. 271.27810-27817(1996)			
CC -I- FUNCTION: TRANSFERS FUCOSE FROM GDP-FUC TO N-LINKED TYPE COMPLEX GLYCOPETIDES. THE PH OPTIMUM IS 7.			
CC -I- CATALYTIC ACTIVITY: GDP-L-FUCOSE + N4-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-MANNOSYL-BETA-N-ACETYL-1,4-D-GLUCOSAMINYL-1,4-N-ACETYL-D-1,2-ALPHA-D-MANNOSYL-1,3-(R(1)-ALPHA-1,6)-BETA-D-MANNOSYL-1,4-BETA-N-ACETYL-D-GLUCOSAMINYL-1,4-(ALPHA-L-FUCOSYL-1,6)-N-ACETYL-D-GLUCOSAMINYL)ASPARAGINE.			
CC -I- PATHWAY: GLYCOSYLATION.			
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND FORM IN TRANS CISTERNAE OF GOLGI.			
CC -I- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.			
CC EMBL: D86723; BAA13157.1; -;			
DR InterPro: IPR001452; SH3.			
DR Pfam: PF00018; SH3; 1.			
DR SMART: SM00326; SH3; 1.			

DR PROSITE: P550002; SH3; 1.  
 KM Transferase; Glycosyltransferase; Transmembrane; signal-anchor;  
 GW Golgi stack.  
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 10 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 26 575 LUMENAL, CATALYTIC (POTENTIAL).  
 SQ SEQUENCE 575 AA; 66229 MW; 0F19DD8C2018F7B CRC64;

Query Match 99.7%; Score 3088; DB 6; Length 575;  
 Best local similarity 99.8%; Pred. No. 4e-225;  
 Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMLILFAWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60  
 DB 1 MRPWGSMRWIMLILFAWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60  
 QY 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEOIENKKOTKNGPKGDHILRRRIE 120  
 DB 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEOIENKKOTKNGPKGDHILRRRIE 120  
 QY 121 NGAKELMFLQSELEKRLKLNKLEGNELORHADEFLLSDLGHNERSIMTDLYYLSQTDGAGDMR 180  
 DB 121 NGAKELMFLQSELEKRLKLNKLEGNELORHADEFLLSDLGHNERSIMTDLYYLSQTDGAGDMR 180  
 QY 181 EKEAKDLTELVRRTIYLLQNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240  
 DB 181 EKEAKDLTELVRRTIYLLQNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240  
 QY 241 LALESHNMWRATGWEIVPRPVSECTDRSGSTGHMSGCVKKNQVVELPIVDSVHPR 300  
 DB 241 LALESHNMWRATGWEIVPRPVSECTDRSGSTGHMSGCVKKNQVVELPIVDSVHPR 300  
 QY 301 PLYPLAVPEDLADRLVRVHGDPVAVMVVSQFVKYLLRPQWLEKEIEATKLGFKHPRV 360  
 DB 301 PLYPLAVPEDLADRLVRVHGDPVAVMVVSQFVKYLLRPQWLEKEIEATKLGFKHPRV 360  
 QY 361 GVHVRRTDKYGAAAHPRPIEYTVHVEEDFQLLARRMQVDKKRVYLATDPPALLKKAATK 420  
 DB 361 GVHVRRTDKYGAAAHPRPIEYTVHVEEDFQLLARRMQVDKKRVYLATDPPALLKKAATK 420  
 QY 421 YPVEYFISDSISMSAGLHNRYTENSIRGVYLLDIHFLSQADFLVCFSSOVCAVAEIMQ 480  
 DB 421 YPVEYFISDSISMSAGLHNRYTENSIRGVYLLDIHFLSQADFLVCFSSOVCAVAEIMQ 480  
 QY 481 ALHPDASANFRSLDDIYFEGGPAHNOAIAYPHQPTGEGIPMEPGDIIGVAGNHMDGYP 540  
 DB 481 ALHPDASANFRSLDDIYFEGGPAHNOAIAYPHQPTGEGIPMEPGDIIGVAGNHMDGYP 540  
 QY 541 KGVNRKLGRTGLPSYKVKREKITVYKPYTPPEADK 575  
 DB 541 KGVNRKLGRTGLPSYKVKREKITVYKPYTPPEADK 575

## RESULT 2

000235 PRELIMINARY: PRT: 575 AA.

AC 000235;  
 DT 01-JUL-1997 (Tremblrel. 04, Created)  
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 GN N-ACYTYL-BETA-D-GLUCOSAMINIDE (ALPHA1-6FUCOSYLTRANSFERASE).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBITaxID=9606;  
 RN 11)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97279058; PubMed=9133635;  
 RA Yanagidani S., Uozumi N., Ihara Y., Miyoshi E., Yamaguchi N.,  
 RA Taniguchi N.,  
 RT Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-  
 glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 fucT) from human

RT gastric cancer MKN45 cells.";  
 RL J. Biochem. 121:626-632(1997).  
 RN 12)  
 RP SEQUENCE OF 278-419 FROM N.A.  
 RA Yamaguchi Y.,  
 RL Submitted (Dec-1997) to the EMBL/GenBank/DBJ databases.  
 RN 13)  
 RP SEQUENCE FROM N.A.

RC TISSUE-WHOLE EMBRYO;  
 RA Cailleau A., Balanzino L., Candelier J.J., Ortol R., Mollicone R.;  
 RT "Differential splice variants of human FUT8 embryonic cDNA."  
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: D89289; BAA19764.1; -;  
 DR EMBL: AF038281; AAB92372.2; -;  
 DR EMBL: AF038280; AAB92372.2; JOINED.  
 DR EMBL: Y17979; CAA76988.1; -;  
 DR EMBL: Y17976; CAA76985.1; -;  
 DR EMBL: Y17977; CAA76986.1; -;  
 DR EMBL: Y17978; CAA76987.1; -;  
 DR InterPro: IP001452; SH3.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART: SM00326; SH3; 1.  
 KM Transferase; Glycosyltransferase.  
 SQ SEQUENCE 575 AA; 66515 MW; 5AE24A93881E18D0 CRC64;

Query Match 95.3%; Score 2949; DB 4; Length 575;  
 Best local similarity 95.5%; Pred. No. 1.3e-214;  
 Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMLILFAWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60  
 DB 1 MRPWGSMRWIMLILFAWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60  
 QY 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEOIENKKOTKNGPKGDHILRRRIE 120  
 DB 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEOIENKKOTKNGPKGDHILRRRIE 120  
 QY 121 NGAKELMFLQSELEKRLKLNKLEGNELORHADEFLLSDLGHNERSIMTDLYYLSQTDGAGDMR 180  
 DB 121 NGAKELMFLQSELEKRLKLNKLEGNELORHADEFLLSDLGHNERSIMTDLYYLSQTDGAGDMR 180  
 QY 181 EKEAKDLTELVRRTIYLLQNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240  
 DB 181 EKEAKDLTELVRRTIYLLQNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240  
 QY 241 LALESHNMWRATGWEIVPRPVSECTDRSGSTGHMSGCVKKNQVVELPIVDSVHPR 300  
 DB 241 LALESHNMWRATGWEIVPRPVSECTDRSGSTGHMSGCVKKNQVVELPIVDSVHPR 300  
 QY 301 PLYPLAVPEDLADRLVRVHGDPVAVMVVSQFVKYLLRPQWLEKEIEATKLGFKHPRV 360  
 DB 301 PLYPLAVPEDLADRLVRVHGDPVAVMVVSQFVKYLLRPQWLEKEIEATKLGFKHPRV 360  
 QY 361 GVHVRRTDKYGAAAHPRPIEYTVHVEEDFQLLARRMQVDKKRVYLATDPPALLKKAATK 420  
 DB 361 GVHVRRTDKYGAAAHPRPIEYTVHVEEDFQLLARRMQVDKKRVYLATDPPALLKKAATK 420  
 QY 421 YPVEYFISDSISMSAGLHNRYTENSIRGVYLLDIHFLSQADFLVCFSSOVCAVAEIMQ 480  
 DB 421 YPVEYFISDSISMSAGLHNRYTENSIRGVYLLDIHFLSQADFLVCFSSOVCAVAEIMQ 480  
 QY 481 ALHPDASANFRSLDDIYFEGGPAHNOAIAYPHQPTGEGIPMEPGDIIGVAGNHMDGYP 540  
 DB 481 ALHPDASANFRSLDDIYFEGGPAHNOAIAYPHQPTGEGIPMEPGDIIGVAGNHMDGYP 540  
 QY 541 KGVNRKLGRTGLPSYKVKREKITVYKPYTPPEADK 575  
 DB 541 KGVNRKLGRTGLPSYKVKREKITVYKPYTPPEADK 575

## RESULT 3

09NOW2 PRELIMINARY: PRT: 575 AA.



AC Q9N0W2: 01-DEC-2000 (TREMBLrel. 15, Created)  
DT 01-DEC-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 6-ALPHA-L-FUCOSYLTRANSFERASE.  
GN FUR8.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RA Jeraud C., Dupuy F., Maftah A., Michalski J.C., Oriol R., Petit J.M.,  
RA Julien R.;  
RT "Ancestral exonic organization of FUR8, the gene encoding the alpha6-  
RT fucosyltransferase reveals successive peptide domains which suggest a  
RT particular 3D core-structure for the alpha6-fucosyltransferase  
RT family."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF247186; AAF65460.1; -.  
DR InterPro: IPR001452; SH3.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00002; SH3; 1.  
KM Transferrase: Glycosyltransferase.  
SQ SEQUENCE 575 AA; 66492 MW; FE04C66B8A5BF540 CRC64;

Query Match 95.1%; Score 2945; DB 6; Length 575;  
Best Local Similarity 95.1%; Pred. No. 2.6e-214;  
Matches 547; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMLILFANGTLLFYIGHLVYRNDHSDHSSRELSTKLAKLERLKOONEDL 60  
DB 1 MRPWGSMRWIMLILFANGTLLFYIGHLVYRNDHSDHSSRELSTKLAKLERLKOONEDL 60  
QY 61 RRMASLSLRIPEDIDOGSPASGRVRLAEQFMKAKQIENYKKQTNKNGKDHIELRRRIE 120  
DB 61 RRMASLSLRIPEDIDOGSPASGRVRLAEQFMKAKQIENYKKQTNKNGKDHIELRRRIE 120  
QY 121 NGAKELMEFFLOSELKLNLEGNELQRNADEFLSDLGHNSRIMTDLVYLSQTDAGDMR 180  
DB 121 NGAKELMEFFLOSELKLNLEGNELQRNADEFLSDLGHNSRIMTDLVYLSQTDAGDMR 180  
QY 181 EKEADDELVQRRITTYLQNPDKSKAKKLVYCNKGGCGQLHNVVYCFPIATGQRT 240  
DB 181 EKEADDELVQRRITTYLQNPDKSKAKKLVYCNKGGCGQLHNVVYCFPIATGQRT 240  
QY 241 LALESHNRYATGMEYFRPVSETCTDRSGSSGTGMSGEVDKNVQVELPIYDSVHR 300  
DB 241 LALESHNRYATGMEYFRPVSETCTDRSGSSGTGMSGEVDKNVQVELPIYDSVHR 300  
QY 301 PLYLAVBEDLADRLVYHGDPAVMWVSQFYKYLIRQPMLEKEIEEATKKLGFKHPYI 360  
DB 301 PLYLAVBEDLADRLVYHGDPAVMWVSQFYKYLIRQPMLEKEIEEATKKLGFKHPYI 360  
QY 361 GVHVARTKVGAEAFHPIEEYTVVEEDFOLLARMQVDDKRVYLATDDPALLKEATK 420  
DB 361 GVHVARTKVGAEAFHPIEEYTVVEEDFOLLARMQVDDKRVYLATDDPALLKEATK 420  
QY 421 YPSYEFISDNTSMAGLHNRYTENS LGVILDIHFLSQADLVCTFSSQCVRAVEIEMQ 480  
DB 421 YPSYEFISDNTSMAGLHNRYTENS LGVILDIHFLSQADLVCTFSSQCVRAVEIEMQ 480  
QY 481 ALHPDASANFSLDIIYFSGPNAHQIAIYPHQRTGEGEIPMEBGDIIIGVAGNHMDGYP 540  
DB 481 ALHPDASANFSLDIIYFSGPNAHQIAIYPHQRTGEGEIPMEBGDIIIGVAGNHMDGYP 540  
QY 541 KGVNKKLGRGTGLPSYKVKREKIEYKPYTPPAEK 575  
DB 541 KGVNKKLGRGTGLPSYKVKREKIEYKPYTPPAEK 575

RESULT 4  
ID Q921U1 PRELIMINARY; PRT: 575 AA.  
AC Q921U1: 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE FUCOSYLTRANSFERASE 8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC010666; AAH10666.1; -.  
KM Transferrase.  
SQ SEQUENCE 575 AA; 66556 MW; 226092A8959B3EB7 CRC64;

Query Match 93.8%; Score 2903; DB 11; Length 575;  
Best Local Similarity 93.4%; Pred. No. 3.9e-211;  
Matches 537; Conservative 14; Mismatches 24; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMLILFANGTLLFYIGHLVYRNDHSDHSSRELSTKLAKLERLKOONEDL 60  
DB 1 MRPWGSMRWIMLILFANGTLLFYIGHLVYRNDHSDHSSRELSTKLAKLERLKOONEDL 60  
QY 61 RRMASLSLRIPEDIDOGSPASGRVRLAEQFMKAKQIENYKKQTNKNGKDHIELRRRIE 120  
DB 61 RRMASLSLRIPEDIDOGSPASGRVRLAEQFMKAKQIENYKKQTNKNGKDHIELRRRIE 120  
QY 121 NGAKELMEFFLOSELKLNLEGNELQRNADEFLSDLGHNSRIMTDLVYLSQTDAGDMR 180  
DB 121 NGAKELMEFFLOSELKLNLEGNELQRNADEFLSDLGHNSRIMTDLVYLSQTDAGDMR 180  
QY 181 EKEADDELVQRRITTYLQNPDKSKAKKLVYCNKGGCGQLHNVVYCFPIATGQRT 240  
DB 181 EKEADDELVQRRITTYLQNPDKSKAKKLVYCNKGGCGQLHNVVYCFPIATGQRT 240  
QY 241 LALESHNRYATGMEYFRPVSETCTDRSGSSGTGMSGEVDKNVQVELPIYDSVHR 300  
DB 241 LALESHNRYATGMEYFRPVSETCTDRSGSSGTGMSGEVDKNVQVELPIYDSVHR 300  
QY 301 PLYLAVBEDLADRLVYHGDPAVMWVSQFYKYLIRQPMLEKEIEEATKKLGFKHPYI 360  
DB 301 PLYLAVBEDLADRLVYHGDPAVMWVSQFYKYLIRQPMLEKEIEEATKKLGFKHPYI 360  
QY 361 GVHVARTKVGAEAFHPIEEYTVVEEDFOLLARMQVDDKRVYLATDDPALLKEATK 420  
DB 361 GVHVARTKVGAEAFHPIEEYTVVEEDFOLLARMQVDDKRVYLATDDPALLKEATK 420  
QY 421 YPSYEFISDNTSMAGLHNRYTENS LGVILDIHFLSQADLVCTFSSQCVRAVEIEMQ 480  
DB 421 YPSYEFISDNTSMAGLHNRYTENS LGVILDIHFLSQADLVCTFSSQCVRAVEIEMQ 480  
QY 481 ALHPDASANFSLDIIYFSGPNAHQIAIYPHQRTGEGEIPMEBGDIIIGVAGNHMDGYP 540  
DB 481 ALHPDASANFSLDIIYFSGPNAHQIAIYPHQRTGEGEIPMEBGDIIIGVAGNHMDGYP 540  
QY 541 KGVNKKLGRGTGLPSYKVKREKIEYKPYTPPAEK 575  
DB 541 KGVNKKLGRGTGLPSYKVKREKIEYKPYTPPAEK 575

RESULT 5  
ID Q9WTS2 PRELIMINARY; PRT: 575 AA.  
AC Q9WTS2: 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ALPHA-L-6-FUCOSYLTRANSFERASE.

CN FUR8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=20358720; PubMed=10902914;  
 RA Hayashi H., Yoneda A., Asada M., Ikekita M., Imamura T.;  
 RT "Molecular cloning of mouse alpha-1,6-fucosyltransferase and  
 expression of its mRNA in the developing cerebrum.";  
 RL DNA Seq. 11:91-96(2000).  
 DR EMBL; AB025198; BAA76392.1;  
 DR MGI; MGI:1858901; Fur8.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART; SM00326; SH3; 1.  
 KW Transferase; Glycosyltransferase.  
 SQ SEQUENCE 575 AA; 66555 MW; 7BE2ED3146E0C45F CRC64;

Query Match 93.4%; Score 2892; DB 11; Length 575;

Best Local Similarity 92.9%; Pred. No. 2,6e-210;

Matches 554; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRPWTGSMWIMLILFAMGTLFLFYIGHLVARDNDSDHSSRELKILAKLERLKQONEDL 60  
 DB 1 MRATGSMWIMLILFAMGTLFLFYIGHLVARDNDSDHSSRELKILAKLERLKQONEDL 60  
 QY 61 RRMAGSLRPEGIIDGPRASGRVRALEEDPRMKEEOTENYKKTQKPGKDHILARRTE 120  
 DB 1 MRATGSMWIMLILFAMGTLFLFYIGHLVARDNDSDHSSRELKILAKLERLKQONEDL 60  
 QY 61 RRMAGSLRPEGIIDGPRASGRVRALEEDPRMKEEOTENYKKTQKPGKDHILARRTE 120  
 DB 61 RRMAGSLRPEGIIDGPRASGRVRALEEDPRMKEEOTENYKKTQKPGKDHILARRTE 120  
 QY 121 NGAKELWFLQSLTKLKLNKGNELORNADEFSLDGNHRSIMTDLUYYSOTDGAQDMR 180  
 DB 121 NGAKELWFLQSLTKLKLNKGNELORNADEFSLDGNHRSIMTDLUYYSOTDGAQDMR 180  
 QY 121 NGAKELWFLQSLTKLKLNKGNELORNADEFSLDGNHRSIMTDLUYYSOTDGAQDMR 180  
 DB 121 NGAKELWFLQSLTKLKLNKGNELORNADEFSLDGNHRSIMTDLUYYSOTDGAQDMR 180  
 QY 181 EKEAKDLTELVRRTIYLLONPKDCSKAKKLVNINCKGCGCOLHHVYVCFMAYGTOPT 240  
 DB 181 EKEAKDLTELVRRTIYLLONPKDCSKAKKLVNINCKGCGCOLHHVYVCFMAYGTOPT 240  
 QY 181 EKEAKDLTELVRRTIYLLONPKDCSKAKKLVNINCKGCGCOLHHVYVCFMAYGTOPT 240  
 DB 181 EKEAKDLTELVRRTIYLLONPKDCSKAKKLVNINCKGCGCOLHHVYVCFMAYGTOPT 240  
 QY 241 LALSHNRKATGGEVTPRVSECTDSSGSTGMSGCVKKNQVVELPYDSVHPR 300  
 DB 241 LALSHNRKATGGEVTPRVSECTDSSGSTGMSGCVKKNQVVELPYDSVHPR 300  
 QY 241 LALSHNRKATGGEVTPRVSECTDSSGSTGMSGCVKKNQVVELPYDSVHPR 300  
 DB 241 LALSHNRKATGGEVTPRVSECTDSSGSTGMSGCVKKNQVVELPYDSVHPR 300  
 QY 301 PPLPLAVPDLADRLVRYVGDPAVMWVSQFYLLIRPOPMLKEIEEATKKGFKHPVI 360  
 DB 301 PPLPLAVPDLADRLVRYVGDPAVMWVSQFYLLIRPOPMLKEIEEATKKGFKHPVI 360  
 QY 301 PPLPLAVPDLADRLVRYVGDPAVMWVSQFYLLIRPOPMLKEIEEATKKGFKHPVI 360  
 DB 301 PPLPLAVPDLADRLVRYVGDPAVMWVSQFYLLIRPOPMLKEIEEATKKGFKHPVI 360  
 QY 361 GVHVRRTDKVGAFAHPRIEETVHVEDFQLLARRMVDKRRVYLATDDPALLKEAKTK 420  
 DB 361 GVHVRRTDKVGAFAHPRIEETVHVEDFQLLARRMVDKRRVYLATDDPALLKEAKTK 420  
 QY 361 GVHVRRTDKVGAFAHPRIEETVHVEDFQLLARRMVDKRRVYLATDDPALLKEAKTK 420  
 DB 361 GVHVRRTDKVGAFAHPRIEETVHVEDFQLLARRMVDKRRVYLATDDPALLKEAKTK 420  
 QY 421 YPSEFTSDNSISWAGLNHRRTENSIRGYIIDLHLSQADFLVCFSSQCVAYEIMQ 480  
 DB 421 YPSEFTSDNSISWAGLNHRRTENSIRGYIIDLHLSQADFLVCFSSQCVAYEIMQ 480  
 QY 421 YPSEFTSDNSISWAGLNHRRTENSIRGYIIDLHLSQADFLVCFSSQCVAYEIMQ 480  
 DB 421 YPSEFTSDNSISWAGLNHRRTENSIRGYIIDLHLSQADFLVCFSSQCVAYEIMQ 480  
 QY 481 ALHPDASANFRSLDDIYFPGCPNAHNOIATYHPQRTGEGIRPEPQDITGVAGNMHMGYP 540  
 DB 481 ALHPDASANFRSLDDIYFPGCPNAHNOIATYHPQRTGEGIRPEPQDITGVAGNMHMGYP 540  
 QY 481 ALHPDASANFRSLDDIYFPGCPNAHNOIATYHPQRTGEGIRPEPQDITGVAGNMHMGYP 540  
 DB 481 ALHPDASANFRSLDDIYFPGCPNAHNOIATYHPQRTGEGIRPEPQDITGVAGNMHMGYP 540  
 QY 541 KGVNRKLGRTGLPSYKVRKEITVYPTPYPEADK 575  
 DB 541 KGVNRKLGRTGLPSYKVRKEITVYPTPYPEADK 575  
 QY 541 KGVNRKLGRTGLPSYKVRKEITVYPTPYPEADK 575  
 DB 541 KGVNRKLGRTGLPSYKVRKEITVYPTPYPEADK 575

RESULT 6 Query Match 46.2%; Score 1431.5; DB 5; Length 619;

Best Local Similarity 44.4%; Pred. No. 7,9e-100;

Matches 270; Conservative 110; Mismatches 179; Indels 49; Gaps 7;

QY 7 SW-RWIMLILFAMGTLFLFYIGHLVARDN-----DHSRRELKILAKLERLKQONEDL 60  
 DB 14 SWARALILFLVAMGILVYVYVLTWTGQQAAGESELNARRISQLOMLEHTRONEEL 73

DE CG2448 PROTEIN (GH11963P).  
 GN CG2448.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Suton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plantkoc C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,  
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Burris J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y, CN BW SP;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF003487; AAF48079.1; -;  
 DR EMBL; AY051451; AAK92875.1; -;  
 DR FLYbase; FBgn0030327; CG2448.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam; PF00018; SH3; 1.  
 DR SMART; SM00326; SH3; 1.  
 SQ SEQUENCE 619 AA; 70205 MW; 691BFD9B56557DE CRC64;

```
OY 61 RRMAGSLRIPGPIIDGPGASGRVAL-----EEOFMKAKEOJENYKK 102
DB 74 KOLIDELMSDO--LDKOSAMKLVQRLENDALNPKLAPVAGPEPESMBESAPADLRGMN 131
OY 103 QTKNGPGK-----DHEILRRRIENGAKELMFLOSELKKLKLML-----EG 142
DB 132 VAEGAPNPLEAGVPDGHGFEPSELEFTRRRQOTIGELIMNFSSSELCKVRKAVAGHAS 191
OY 143 NELORHADEFSLDGHHERSIMTDLIYLSQTDGAGDMREKAKDITELVQRITTYLQNP 202
DB 192 ADLESISINQVLQGAENHRSLSLDMERKOSDGEAMRNHKEARDLSDVQORLHLQDPS 251
OY 203 DCSKAKKLVACNINKGCGCOLHNHYVCFMAYGQPTLALSHNMWRATGEMETVFPV 262
DB 252 DCQNRKRLVCKLKNKCGGCOLHNHYVCFMAYATERTLLKSKGRMYHKGMEVEFPV 311
OY 263 SETCTDRSGSSGTGHWGSEVKDKKNQVVELPIVDSVHPRPYPLAVPEDLADRLVRHGD 322
DB 312 SNSCHDACTANTYNNPK---PNTQVLYPLIIDSIMPRPYPLAVPEDLAPRLKRLHGD 368
OY 323 PAVKVVSGPVKYLIRPQWLEKEIEATKKLGFKHPIYGVHVRTPDKVGAEAHFPIEBY 382
DB 369 PIVWVVGQFLKYLIRPQPTTRDFLTSGMRNLGWERPIVGVHVRTPDKVGTAAACHSVEEY 428
OY 383 TVHVEDEFLARMQVQDKKRYLATDDPDLKLEAKTKYPSYEFTSDMSISAGLHNR 442
DB 429 MTYVEDYRTLELVNGSTVARRLEFLASDDQVYEEARKKIPQOITIGDEPVARMAVSSTRY 488
OY 443 TENSRLGYLDIHFLSQADFLVCFPSSQVCRVAYEIMQALHPDASANFRSLDDIYFSGP 502
DB 489 TDTALNGIILDIHLMSDLHLCFTSSQVCRVAYEIMQTMVDAHNRKSLDDIYVYGO 548
OY 503 NAHNOIATYHPQRTPEGEIPEBPDIIGVAGNHMDGCRVKNRKLGRGLPSYVYREKI 562
DB 549 NAHNRVYIAHPRTHEDLQLRVGDLYSAGNHMDGNSKKNRTNQGCLFSPFVEEKV 608
OY 563 ETVKYPTY 570
DB 609 DTAKLPLY 616

RESULT 7
O9P2U6 PRELIMINARY: PRT: 276 AA.
AC O9P2U6:
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DB 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DB 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALPHAI.6 FUCOSYLTRANSFERASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20275614; PubMed=10814706;
RA Yamaguchi Y., Ikeda Y., Takahashi T., Ihara H., Tanaka T., Sasao C.,
RA Uozumi N., Yanagidani S., Inoue S., Fujii J., Taniguchi N.;
RT "Genomic structure and promoter analysis of the human alpha1,6-
RT fucosyltransferase gene (FUT8).";
RL Glycobiology 10:637-643(2000).
DB EMBL: AB032572; BAA92858.1;
DB EMBL: AB032568; BAA92858.1; JOINED.
DB EMBL: AB032569; BAA92858.1; JOINED.
DB EMBL: AB032570; BAA92858.1; JOINED.
DB EMBL: AB032571; BAA92858.1; JOINED.
KW Transferase; Glycosyltransferase.
FT NON-TER 276
FT SEQUENCE 276 AA; 32182 MW; 18BRC2AC59152E61 CRC64;

Query Match 45.9%; Score 1420; DB 4; Length 276;
Best Local Similarity 95.7%; Pred. No. 1.9e-99;
```

```
Matches 264; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
OY 1 MRPTGSRWIMLILFANGTLFTYIGHLVNDNDHSDSSSELKYLAKLERLKOONDL 60
DB 1 MRPTGSRWIMLILFANGTLFTYIGHLVNDNDHSDSSSELKYLAKLERLKOONDL 60
OY 61 RRMAGSLRIPGPIIDGPGASGRVALLEOFPMAKKEOJENYKOTKNGPGKHETLRRRIE 120
DB 61 RRMAGSLRIPGPIIDGPGASGRVALLEOFPMAKKEOJENYKOTKNGPGKHETLRRRIE 120
OY 121 NGAKELMFLOSELKLEKLEGNELQRNADEFSLDGHHERSIMTDLIYLSQTDGAGDMR 180
DB 121 NGAKELMFLOSELKLEKLEGNELQRNADEFSLDGHHERSIMTDLIYLSQTDGAGDMR 180
OY 181 EKEAKDTELVRRTTYLQNPDKCSKAKKLVACNINKGCGCOLHNHYVCFMAYGTORT 240
DB 181 EKEAKDTELVRRTTYLQNPDKCSKAKKLVACNINKGCGCOLHNHYVCFMAYGTORT 240
OY 241 LALSHNMWRATGEMETVFPVSETCTDRSGSSGTGH 276
DB 241 LALSHNMWRATGEMETVFPVSETCTDRSGISGTGH 276

RESULT 8
O9BYC6 PRELIMINARY: PRT: 246 AA.
AC O9BYC6:
DB 01-JUN-2001 (TREMBLrel. 17, Created)
DB 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DB 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHAI.6-FUCOSYLTRANSFERASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP TISSUE=RETINA;
RC Yamaguchi Y., Ikeda Y., Ookawara T., Suzuki K., Taniguchi N.;
RA "The splicing variant of the human alpha 1,6-fucosyltransferase.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DB EMBL: AB049740; BAA40929.1;
KW Transferase; Glycosyltransferase.
FT NON-TER 1
FT SEQUENCE 246 AA; 28310 MW; BA69CBD960458E14 CRC64;

Query Match 35.8%; Score 1109; DB 4; Length 246;
Best Local Similarity 94.9%; Pred. No. 5.2e-76;
Matches 206; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
OY 63 MAGSLRIPGPIIDGPGASGRVALLEOFPMAKKEOJENYKOTKNGPGKHETLRRRIENG 122
DB 1 MAESLRIPGPIIDGPGASGRVALLEOFPMAKKEOJENYKOTKNGPGKHETLRRRIENG 60
OY 123 AKELMFLOSELKLEKLEGNELQRNADEFSLDGHHERSIMTDLIYLSQTDGAGDMREK 182
DB 61 AKELMFLOSELKLEKLEGNELQRNADEFSLDGHHERSIMTDLIYLSQTDGAGDMREK 120
OY 183 EAKDTELVRRTTYLQNPDKCSKAKKLVACNINKGCGCOLHNHYVCFMAYGTORTLA 242
DB 121 EAKDTELVRRTTYLQNPDKCSKAKKLVACNINKGCGCOLHNHYVCFMAYGTORTLI 180
OY 243 LESHNMWRATGEMETVFPVSETCTDRSGSSGTGHSG 279
DB 181 LESHNMWRATGEMETVFPVSETCTDRSGISGTGHSG 217

RESULT 9
O16882 PRELIMINARY: PRT: 818 AA.
AC O16882:
DB 01-JAN-1998 (TREMBLrel. 05, Created)
DB 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
```



```
Qy 421 YPSYEFISDINSAGLHNRTENSLRGVILDIHFLSQADFLVCTFSSQ 470
Db 1 YPNEEFISDINSAGLHNRTENSLRGVILDIHFLSQADFLVCTFSSQ 50

RESULT 12
09SAAS PRELIMINARY; PRT; 1025 AA.
AC 09SAAS;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F25C20.13 PROTEIN.
GN F25C20.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremenetskaya I., Luros J., Ngan I., Liu A.,
RA Gonzalez A., Alifafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Hultzer L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,
RA "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007296; AAD30251.1; -.
SQ SEQUENCE 1025 AA; 116411 MW; 08C952A2032BA1E4 CRC64;

Query Match 4.5%; Score 140; DB 10; Length 1025;
Best Local Similarity 17.8%; Pred. NO. 0.072;
Matches 110; Conservative 80; Mismatches 193; Indels 234; Gaps 27;

Qy 3 PWTGSRMRLMILFAMGILFF-YI-----GGLVPRNDSDHSSRELKTLAKLERLK 54
Db 360 PKSGNW-WFAEYVPGGALVDWVADGPRGAFLYDNGGIDFHALVPQKLPRELYWLE 418

Qy 55 QONEDLRMAGSLRIPEGRITDQSPASGRVRLAEQFMKAK-EQIENYKQTKNGPKDHE 113
Db 419 EENMIFRLQED-----RRLKEVMAKMKETARLKAETKERTLKKEFL 461

Qy 114 ILRR-----RIENGAKELMFF-----LOSELKKL 137
Db 462 LSQKDVVYTERPLEIQAGNPVTVLYNPANTVLYNGKPEVWVRGSGFNMTHTRLGPRQKMEA 521

Qy 138 KNLEGENEIORNA---DEFLISLGHHERS-----162
Db 522 TDDESSHVKTAKVPLDVAWMDVFSEKEDGIFDNKKGLDHLRVGSGISKEPRLIYH 581

Qy 163 IMTDLVYLSQTDGAGDMEKEAKDLTELVRORITVLYQNPKDCSK--AKKLVCN----- 213
Db 582 IAVENAPLAKVGGGLDVYTSLSRAVQEL-NHNVDIYVFPKYDDIKINFPKXDLQFNSTYHMG 640

Qy 214 -----INKGCGYGO-----LHHVVCENIAYGTORT 240
Db 641 GTEIKVMHGKVEGLSVYFLDPONGLFQRCGVYGCADADAGREGFCHALAEFLQCGFHPD 700

Qy 241 LALESHNNRYATGGEWETVPRVSEFCTDRSSSTGHWSEGVNDKAVQ--VVEFLVDSVH 298
Db 701 I-LHCHDWSAPVSW--LF-----KDHYTQGLIKITRIVFTIH 735

Qy 299 -----PRPYL-PLAVPEDLADRLVAVHG-----DPAVW--WVS 329
Db 736 NLEFGANNIGKAMTFADKATTVSPYIAKEVAGNSVISAHLVYFHHIINCIDIDIDIDPIND 795
```

```
Qy 330 QPVKYLIPQPMLEKEI-----EATKKLGK---HPYIGVHVNRRTDKVGAENAFHP 378
Db 796 NFI-----PVPTSENVVEGKRAAKEELQNLGKLSADPFPVVGITTRJTHOKGIHLJKHA 850

Qy 379 I-----EEYTVHEEDFOLLARRMOV---DKKRYVLATDDPALMLEAKTKY 421
Db 851 IWRTLERNQGVYLLGSADPPRIQNDVFVLANQHLSSHCDRRRLVLTIDEPL-----SHLIY 906

Qy 422 PSEYEFISDINSAGL 438
Db 907 AGADFLVPSIFPEPCGL 923

RESULT 13
099UR5 PRELIMINARY; PRT; 1057 AA.
AC 099UR5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CARBAMOYL-PHOSPHATE SYNTHASE LARGE CHAIN.
GN PYRAB OR SA1046 OR SAV1203.
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 153878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mizutani-Ui Y., Takehashi N.K., Sawano T., Inoue R.-T., Kato C.,
RA Sekitani Z.K., Hiraoka H., Kohara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiratazu K.;
RA "Whole genome sequencing of methicillin-resistant staphylococcus
RL aureus."
RT Lancet 357:1225-1240(2001).
DR EMBL: AP003132; BA842298.1; -.
DR EMBL: AP003361; BAB57365.1; -.
DR HSSP; P00968; ICS0.
DR InterPro; IPR000901; CPSase.
DR InterPro; IPR000169; Thiolprot_act_site.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR PRINTS; PR00098; CPSase.
DR PROSITE; PS00866; CPSASE.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_2.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

Query Match 4.0%; Score 123.5; DB 16; Length 1057;
Best Local Similarity 18.9%; Pred. NO. 1.3;
Matches 115; Conservative 85; Mismatches 221; Indels 189; Gaps 25;

Qy 43 LSKILAKLERLKQONEDLRMAGS-----LRIPGEPID-----OGP 78
Db 318 IAKLAKIYAVGLTLEMLNPITGSTYAFPTLDYVSKIPRPFPDKREKREGELGTQMK 377

Qy 79 ASGRVAL-----EEOFMKAKEQIEN--YKQTKNGPGKDHELKRRINGAKELMFFLQS 132
Db 378 ATGEVMAIGRYEESILKAIRSLFEGVNHGLPNCESFDLDYIKERISHODERLFFTGE 437

Qy 133 ELKTKLNEGNELOHNADEF-----LSDLGHERSLMPTDLVYLSQTDGAGDMEKEAK 185
```

```
Db 438 AIRGTTLEIEHNMQIDYFFELHKNQIIDEHQLKEHOGDELYLKAYDYG-FSDKITA 496
QY 186 DLTELQORRTTYLQNPDKDSKAKKLVNINKKGGYGCQLHHVYVCFMIAVGTQRTLALS 245
Db 497 HRENMTREEVYQLRMENDIKPVYKMY----- 522
QY 246 HNMRYATGMEVFRPVSETCTDRSGSGHMSGEVYKKNVQVY---ELPIVDSVHPRP 301
Db 523 -----DTCAEFESSTPYYGYGTETENESIIVTDKEKILYLGSG----- 560
QY 302 PLYPLAVPEDLADRLVRYVHGDPVMMVSO--FYKYLIRPQ-----WLEKEL 346
Db 561 ---PIRIGGVEEDYATVH--AVMAIQAGYEAITVNNNPETVSTPFSISDKLYFEPLT 614
QY 347 EEAATKLGKFNPIYGVHVRRTDKVGAEAHFHPEEYTVH-----VEEDF 390
Db 615 EEDVYNNIINEKRGVYV---QFGGQTAINLADKLAKHGVKILGTSLENLNAEDRKKE 670
QY 391 QLLARMQYDKKRVYLATD-DPALKEAKTKY---PSYEF-----ISDNSISWSAGL 438
Db 671 EALLRKINVPQPGKSAATSEPALANAAEIGYPVVRPSYVLGRAMEIYDN---DKEL 726
QY 439 HNNYTE---NSLRGYILDIHFLSODFLYCTPSSQVCRVAYEIMALHPDASANFRSD 494
Db 727 ENMTQAVKASPEHPVLD-RYLTGKEIEVDALCDGETVYIIPGIMHIER----- 775
QY 495 DIYFEGGNANHOIATYHPQRTGEGE-----IPMEPG-DITGVAGNHW---DG-YP 540
Db 776 ----ACVHSGDSIAVYPPQTLTDEDLATLEDTITKLGLNITIGLNIQFYAHNGVYV 830
QY 541 KGVNRKLGRT 550
Db 831 LEVNPRESSRT 840

RESULT 14
Q9SRD5 PRELIMINARY: PRT: 1871 AA.
ID 09SRD5:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE HEAT SHOCK PROTEIN. 53413-59028.
GN F28016.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maitl R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;
RA Arabidopsis thaliana chromosome 1 BAC F28016 genomic sequence.;
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC010718; AAF04452.1;
DR InterPro: IPR002068; CrystalLin_HSP20.
DR Pfam: PR00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Heat shock.
SQ SEQUENCE 1871 AA; 216899 MW; 52300C21F3DBF5A1 CRC64;
```

Query Match 4 0%; Score 123; DB 10; Length 1871;  
Best Local Similarity 25.0%; Pred. No. 3.3;  
Matches 47; Conservative 40; Mismatches 73; Indels 28; Gaps 9;

```
QY 24 YIGGHLVRNDSDHSSRELK-ILAKLERLQONEDLRMGSLRIPGPIDQGPASGR 82
Db 1065 HNGG-----EDNHDHKEEOKENVIKAE-LNTEEDSFKKVEIEKQDNGELKRSVQAK 1118
QY 83 VRALEE-----QFYKAKEQIENYKKQTKNG-----PGKDHEI--LRRIENGAKELWFF 129
```

```
Db 1119 RQTEEKDKTRAMEKNEVER-RKQTKDGLKLRGCEDEPREGHERRGEEDRIEEL--- 1174
QY 130 LOSEL-----KRLKNLEGNLQRNADEFLSDLGHNEHSIMTDLYYLSQDTGACADMKEAK 185
Db 1175 VETESIDHKEKKKDEDEYILRSQDTGKVLDGERERRSKORIKHSVEDEIDQDEDAE 1234
QY 186 DLTELQOR 193
Db 1235 EAAAVYSR 1242

RESULT 15
Q921B9 PRELIMINARY: PRT: 1200 AA.
ID 0921B9:
AC 0921B9:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NUCLEAR PORE COMPLEX-ASSOCIATED PROTEIN TPR (FRAGMENT).
GN TPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/OLA;
RA Sandblad L., Hunziker A., Cordes V.C.;
RT "Evolutionarily conserved mouse tpr is a single-copy gene located on
RT chromosome 1.";
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ298076; CAC40701.1;
DR NON_TER 1200
SQ SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;
```

Query Match 3.9%; Score 122; DB 11; Length 1200;  
Best Local Similarity 21.1%; Pred. No. 2.1;  
Matches 101; Conservative 53; Mismatches 199; Indels 126; Gaps 17;

```
QY 23 FYIGGHLVRNDSDHSSRELKILAKLERLQONEDLRMGSLRIPGPIDQGPASGR 82
Db 55 FIEEKLRSQSORLVTETRECONLRLEKLNQVVLTEKRELE-----TAADR 105
QY 83 VRALEEFKAKEQIENYKKQTKNGPKDHLRRRIENGAKELWFFQSELKTL----- 137
Db 106 NIGIQSOFRAHELEAEKRDLL--RTERLSQVE-----YLIEDVKRLNEKXL 153
QY 138 -KNLSENLQRNADEFLSDLGHNEHSIMTDLYYLSQDTGAGDREKAKDLTELQORIT 196
Db 154 ESNTTKGELQLKDE-----LQASDAVAVYRERKRLRQEKELLNONS 195
QY 197 YLQNPDKSKAKKLVNINKG---CGYGCQLHHVYVCFMIAVGTQRTLALESHNNRYATG 253
Db 196 WNLTELKTKTDELLALGRKGNELIELKCNLEN-----KKEVNLREQMGKLTIS 246
QY 254 GMEVFRPVSETCTDRSGSGTGWSGEVYKKNVQVVELPI-----VDSVPRPPYLP 306
Db 247 N-EHLQKHNEDDLTKLKEKKEEQQASMEKEKFHNLNHIKLSLYSADDSAKKSNELTR 305
QY 307 AVEP-----DLADRVLRYVHGDPVMMVWSQFYKYLIRPQVLEKIEEATYKL 353
Db 306 AVDELKHLKLEAGEANKTIQDLHLQVEESK-----DQMEKEMLEKIGKLEKELNANDLL 360
QY 354 GKFNPIYGVHVRRTDKVGA-----EAANRP-----IEEYTVHVEEDQL 392
Db 361 S-----ATRKKGAILSEBELAASPTAAAVAKIVKGMKLTLEYLYVETQDQL 409
QY 393 LARRMQYDKKRVYLATDDPALKEAKTKY-----SYEISDNSISWSAGLHNNYTE 444
Db 410 LLEKQ--ENKRIKTYLDE--YKVEYKAPILKROREETERQKAVASLSAKLEQAMKE 464
```

Wed Nov 6 14:29:26 2002

us-09-839-136-2.rspt

Search completed: October 31, 2002, 13:09:39  
Job time : 38 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:00 : Search time 14 Seconds

(without alignments)  
1590.267 Million cell updates/sec

Title: US-09-839-136-2

Perfect score: 3096  
Sequence: 1 MRPWGSMRWIMILFAMGT.....YKVRKIEYKPYTPPEADK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122.5	4.0	1427	1 REST_HUMAN	P30622 homo sapien
2	117.5	3.8	2349	1 TPR_HUMAN	P12270 homo sapien
3	117	3.8	1938	1 MYSD_CAEEL	P02567 caenorhabdi
4	114.5	3.7	845	1 SCPL_MOUSE	P06053 mesocricetu
5	112	3.6	352	1 SH32_MOUSE	Q62420 mus musculu
6	111.5	3.6	978	1 RA50_AQUAE	Q67124 aquilex aeo
7	111.5	3.6	1071	1 CARB_BACSU	P25994 bacillus su
8	110.5	3.6	1074	1 PLD1_HUMAN	Q13393 homo sapien
9	110	3.6	976	1 SCPL_HUMAN	Q15431 homo sapien
10	110	3.6	993	1 SCPL_MOUSE	Q62209 mus musculu
11	110	3.6	3210	1 CENF_HUMAN	P49454 homo sapien
12	109.5	3.5	480	1 STEA_METUA	Q57911 methanococc
13	109	3.5	214	1 OSF1_HUMAN	Q92882 homo sapien
14	109	3.5	324	1 NOD2_BRAJA	Q45271 bradyrhizob
15	108.5	3.5	1014	1 NEBL_HUMAN	Q76041 homo sapien
16	106.5	3.4	900	1 MANB_CAEEL	Q93324 caenorhabdi
17	106	3.4	352	1 SH32_HUMAN	Q99962 homo sapien
18	106	3.4	1969	1 MYSA_CAEEL	P12844 caenorhabdi
19	105.5	3.4	646	1 HS70_THEHL	P16019 theilleria a
20	105.5	3.4	1961	1 MYH9_RAT	Q68812 rattus norv
21	105	3.4	673	1 LOX5_HUMAN	P09917 homo sapien
22	105	3.4	880	1 RA50_HUMAN	Q9uzc8 pyrococcus
23	105	3.4	919	1 RA50_AERPE	Q9y121 aetopyrum p
24	103.5	3.3	1966	1 MYSB_CAEEL	P02566 caenorhabdi
25	103.5	3.3	1976	1 MYHA_BOVIN	Q27991 bos taurus
26	103	3.3	887	1 VAV5_SCHPO	Q10213 schizosacch
27	103	3.3	1230	1 UGSA_SOLTRU	Q43846 solanum tub
28	102.5	3.3	1960	1 MYH9_HUMAN	P35579 homo sapien
29	102	3.3	215	1 OSF1_MOUSE	Q62422 mus musculu
30	102	3.3	519	1 ARH5_HUMAN	P12774 homo sapien
31	102	3.3	733	1 MOTB_STRCM	Q05065 streptomyce
32	101.5	3.3	529	1 VSM6_TRYBB	P26334 trypanosoma
33	101.5	3.3	723	1 MY5B_MOUSE	P21271 mus musculu

34	101.5	3.3	997	1 SCPL_RAT	Q03410 rattus norv
35	101	3.3	545	1 CIP4_HUMAN	Q15642 homo sapien
36	101	3.3	3685	1 DMD_HUMAN	P11532 homo sapien
37	100.5	3.2	362	1 RPL_AQUAE	Q67032 aquilex aeo
38	100.5	3.2	1004	1 CARE_HUMAN	Q9bxi6 homo sapien
39	100.5	3.2	1074	1 PLD1_MOUSE	Q92280 mus musculu
40	100.5	3.2	1755	1 PEPL_MOUSE	Q9r269 mus musculu
41	100	3.2	1526	1 MYS2_SCHPO	Q9us16 schizosacch
42	100	3.2	1675	1 POL_RTBVP	P27502 rice tungro
43	100	3.2	1920	1 PCNT_MOUSE	P48725 mus musculu
44	100	3.2	1938	1 MYH4_RABIT	Q28641 oryctolagus
45	99.5	3.2	864	1 RA50_SULSO	Q97wh0 sulfolobus

## ALIGNMENTS

RESULT 1	ID	REST_HUMAN	STANDARD:	PRT: 1427 AA.
AC	P30622:			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-Sternberg intermediate filament associated protein).			
GN	RSN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Plimates; Carnivora; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Peripheral blood monocytes;			
RX	MEDLINE=92289675; PubMed=1600942;			
RA	Bilbe G., Delabie J., Brueggem J., Richner H., Asselbergs F.A.M.,			
RA	Cerletti N., Sorg C., Odink K., Tarsay L., Miesendanger W.,			
RA	de Wolf-Peters C., Shipman R.,			
RT	"Restin: a novel intermediate filament-associated protein highly			
RT	expressed in the Reed-Sternberg cells of Hodgkin's disease.";			
RL	EMBO J. 11:2103-2113(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92405160; PubMed=1356075;			
RA	Pierre P., Scheel J., Rickard J.E., Kreis T.E.;			
RT	"CLIP-170 links endocytic vesicles to microtubules.";			
RL	Cell 70:887-900(1992).			
CC	- FUNCTION: SEEMS TO BE AN INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.			
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE			
CC	CYTOSKELETON.			
CC	- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A			
CC	SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERBERG CELLS			
CC	OF HODGKIN'S DISEASE.			
CC	- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL: X64838; CAA46050.1; -			
DR	EMBL: M97501; AAA35693.1; -			
DR	PIR: S22695; S22695.			
DR	MTM: 179838; -			
DR	InterPro: IPR000938; CAP-Gly.			
DR	InterPro: IPR001878; Znf_CCHC.			
DR	PFam: PF01302; CAP-Gly; 2.			
DR	SMART: SM00343; Znf_C2HC; 1.			

```

DR PROSITE: PS00845; CAP-GLY-1, 2.
DR PROSITE: PS50245; CAP-GLY-2, 2.
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 232 274 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1069 1069 D -> E (IN REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD942544E8 CRC64;

Query Match
Best Local Similarity 4.0%; Score 122.5; DB 1; Length 1427;
Matches 104; Conservative 18.3%; Pred. No. 1;
Mismatches 100; Indels 196; Gaps 22;

OY 40 SRLSLIYLALEKLEKQONEDLRMGSLRIPGPIDQGPASGRVRLDE----- 88
DB 588 SKENESIKSLKSLHANKNSNDV-----IALMKSKLETAISHQ-QAMEELKVSFSKGLGT 640
OY 89 ---QFMKAKQIE---NYKKOTKN-----GRGDHILRRRIENGAKELMFPIQ 131
DB 641 ETAEFEALKTQLEKMLDIOHEIENLONODSRAAHAKEMALRAKMLKVIKEKENSLE 700
OY 132 S-----ELKKLKNLEG--NELQHNADFLSLDGHNR 161
DB 701 AISKLDKADQHLVEMEDTLNKLQAEIKVKELEVLAQAKCNQGTVIDNFTSQLATTE 760
OY 162 SINTDLIYLSQTDGADMDREKAKDITELVQRIITYLONPK--DCKAKKL----- 210
DB 761 KLL-DLDALRKASSEKSEKKLRQOLEAKQIKHLEIKNESSKASSITRELOGREL 819
OY 211 -VCNINMGCGYGGQHLNHYVCFMIAVGTOFTALLESNNRYATGMEYVPRVSECTDR 269
DB 820 KLTNLDENLSEVSQVETL-----EKELOILKEKFAEASEAVORSQOET----- 866
OY 270 SGSSTGHWGEVKKVQVVELPIVDSVHPRPYRLP-----AVBEDLALVRAHGDP 323
DB 867 -----VNKLHQKEQFMNLSLDEKLRENLDMEAKFREK- 901
OY 324 AVWWSQFVKYLIRPQPMLEKEIEEATK-----KLGKRRHYGVGHVNRITDK 369
DB 902 -----DEREQDLIKAREKLENDIAELIMKMSGDSSQLTKMDELRLKERDVELODKLR 956
OY 370 VGEAAFA--HPIREYVHVEDFOFLARRQVQKKRVYLATDPPALKEAKTKYPSYEFT 427
DB 957 ANENASFLQKSIDMTYVKAQSQOEAQKHEEKELERKLSL--LEKKMETSHNOCQEL 1014
OY 428 S---DNSISMSAGLHNR-----TENSILR-----GVILDIHFL-SQADEFLVCT 466
DB 1015 KANERATSETTKKHEEILQNLQKTLTDREDKLGAKEKMSGLOLELELRKQAD---KA 1071
OY 467 FSSQVCRVAVETIMQALHPDASANFRSLDD 495
DB 1072 KAAQTAEADAMQIMEQMTKETETLASLED 1100

RESULT 2
TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (rel. 12, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-93064711; PubMed-1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE-95096166; PubMed-7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "Tpr, a large coiled coil protein whose amino terminus is involved in
activation of oncogenic kinases, is localized to the cytoplasmic
surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE-88262257; PubMed-3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -I- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
COMPONENTS, INCLUDING P62.
CC -I- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
BRAIN. LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -I- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
OR RAF GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X66397; CAA47021.1; -
DR EMBL: Y00672; CAA68681.1; -
DR PIR: S00928; S00928.
DR MIM: 189940; -
KW Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; AFDD6885CEDCA9EF CRC64;

Query Match
Best Local Similarity 3.8%; Score 117.5; DB 1; Length 2349;
Matches 96; Conservative 21.3%; Pred. No. 4.4; Mismatches 184; Indels 119; Gaps 16;

OY 23 FYVGHVLDNDHDSISRELKILAKLEKQONEDLRMGSLRIPGPIDQGPASGR 82
DB 55 FELEKRLSHQGLVETREKQSLRELEKLNQKALTEKKKELEI-----AQR 105
OY 83 VRALAEQFMKAKQIEENYKKOTKNQNGFKDHEILRRRIENGAKELWFFLOSELK 137

```

```
Db 106 NIAIQTSTTRKEELEAKRD-----LRTNRLELQEL-EYLTEDVKRLNEKX 153
Oy 138 -KNEGNELQHADEBLSDLCHHERSIWFDLYISQTDGADWREKAKDLTELQVRIT 196
Db 154 ESNNTTKGLQKLDE-----LQASDVSKYREKRLLEQKELLHSQNT 195
Oy 197 YLQPKDSCAKKLVNCKNG---CGYGCQLHHVYVCPMIAVGTORTLALSHMNRVGTG 253
Db 196 WLNTLTKTKTDELLALGREGNEELLEKCNLEN-----KKEEVSRLQDMNGLKTS 246
Oy 254 GMEVTFRPVSETCTDRSGSTGHSGEYKDKNVQVELPI-----VDSVHPRPYPL 306
Db 247 N-EHLQKVEVELLTKLKAKQOASMEKPHNELNAHKLSTLYKSADDESKSNELTR 305
Oy 307 AVPE-----DLADRLVYHGDPAVWVVSQPVKTLIRPOPLEKEIEEATKKL 353
Db 306 AVEELHKLKLEAGKANKAIDHLLLEEQSK-----DQMEKEMLEKIGRLKELEMANDL 360
Oy 354 GFKHPIVGVHVRITDKVGA-----EAAFH-----IREYTVHVEDPQL 392
Db 361 S-----ATRKKAIIISEELAMSPMAAIVAKIVPKMKLTLYNAVYETDOL 409
Oy 393 LARRMOYDKKRVYLATDPPALLKEAKTKYP 422
Db 410 LLEKL--ENKRINKYLD--IVKEVEAKAP 435

RESULT 3
MYSD_CAEEL STANDARD: PRT: 1938 AA.
AC P02567: 019674:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain D (MHC D).
GN MYO-1 OR R06C7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP STRAIN-BRISTOL N2.
RC MEDLINE=89178677; PubMed=2926820;
RX MEDLINE=89178677; PubMed=2926820;
RA D1db N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
heavy chain gene family."
RL J. Mol. Biol. 205:603-613(1989).
RN 12
RP SEQUENCE OF 34-1795 FROM N.A.
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
myosin heavy chain gene are not separated by introns."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN 13
RP SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
RX MEDLINE=85201409; PubMed=3888374;
RA Karn J., D1db N.J., Miller D.M.;
RT "Cloning nematode myosin genes."
RL Cell Muscle Motil. 6:185-237(1985).
RN 14
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gardner A., McMurtry A.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATOR LIGHT CHAIN SUBUNITS (RLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
```

```
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C. ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-slb.ch/announce/
CC or send an email to license@1sb-slb.ch).
```

DR	EMBL	X08065	CAA30854.1	-	
DR	EMBL	M37232	AAA28119.1	-	
DR	EMBL	M37234	AAA28120.1	-	
DR	EMBL	Z71266	CAA95848.1	-	
DR	EMBL	Z71261	CAA95848.1	JOINED.	
DR	EMBL	Z71261	CAA95806.1	-	
DR	EMBL	Z71266	CAA95806.1	JOINED.	
DR	PIR	S02772	MMKMI.		
DR	HSSP	P08799	IMND.		
DR	WormPep	R06C7.10	CE06253.		
DR	InterPro	IPR004009	Myosin_N.		
DR	InterPro	IPR002928	Myosin_Tail.		
DR	InterPro	IPR001609	myosin_head.		
DR	Pfam	PF00063	myosin_head.1.		
DR	Pfam	PF02736	Myosin_N.1.		
DR	Pfam	PF01576	Myosin_Tail.1.		
DR	PRINTS	PR00193	MYOSINHEAVY.		
DR	ProDom	PD000355	MYOSINHEAD.1.		
DR	SMART	SM00242	MYSC.1.		
KW	Myosin	Muscle protein; Coiled coil; Thick filament; Actin-binding;			
KW	ATP-binding	Methylation; Alkylation; MYOSIN HEAD-LIKE.			
FT	DOMAIN	1	845		
FT	DOMAIN	846	1938		
FT	DOMAIN	1171	1938		ALPHA-HELICAL TAILPIECE (SHORT S2).
FT	DOMAIN	846	1938		LIGHT MEROMYOSIN (LM).
FT	NP_BIND	177	184		COILED COIL (POTENTIAL).
FT	DOMAIN	660	682		ATP (BY SIMILARITY).
FT	DOMAIN	764	778		ACTIN-BINDING.
FT	MOD_RES	128	128		ACTIN-BINDING.
FT	MOD_RES	700	700		METHYLATION (SH-1).
FT	MOD_RES	710	710		ALKYLATION (SH-2).
FT	MOD_RES	94	94		ALKYLATION (SH-2).
FT	CONFLICT	98	98		F -> E (IN REF. 2).
FT	CONFLICT	377	377		A -> R (IN REF. 2).
FT	CONFLICT	389	390		V -> D (IN REF. 4).
FT	CONFLICT	391	391		DV -> GD (IN REF. 2).
FT	CONFLICT	408	408		V -> D (IN REF. 4).
FT	CONFLICT	474	474		W -> N (IN REF. 2).
FT	CONFLICT	577	577		Q -> G (IN REF. 2).
FT	CONFLICT	681	681		L -> F (IN REF. 4).
FT	CONFLICT	1373	1373		I -> N (IN REF. 4).
FT	CONFLICT	1659	1659		S -> D (IN REF. 2).
FT	CONFLICT	1659	1659		E -> Q (IN REF. 3).
SO	SEQUENCE	1938 AA;	223255 MW;	387399C8F63AACF4	CRC64;

Query Match 3.8%; Score 117; DB 1; Length 1938;  
Best Local Similarity 20.5%; Pred. No. 3.7;  
Matches 97; Conservative 70; Mismatches 139; Indels 168; Gaps 21;

```
Oy 36 SDHSSRELSTKLAKLERKQONEDLR---MAGSLRIPEPIDOGPASGR-----V 83
Db 1481 SRNTSTVEFKLRSMINDSEQIETLRRENKIFSGEIRINDINQIOG---GRTYQEVHKSV 1537
```

QY 84 RALEOFMAKEOIEYKQTKNG-PKDHETLRRIENGAKELMFLSELKLNLEQ 142  
 DB 1538 RRLGE-----KDLQNALDEAALAEESKVLAR-----LQEVQGINSEIE 1580  
 QY 143 NEIQRADEFLSLGHNH---SIMTLYLSOTDAGMREKADLTLYVORITVLO 199  
 DB 1581 KRIOEKEEFEENRKNHROSTIQASL-----ETEAESKALRAK----- 1622  
 QY 200 NPIRDSKAKKLVNKGCGGCGGLHHVYCFMIAVGTOFTLESGINMY--ATGGMET 257  
 DB 1623 -----KLEFDIN-----OLEILDHANANVDAQNLKK 1652  
 QY 258 VFRVSECTDRSGSSGTGHSGEVKDMVVELPIVDSVHPRPYL-----PLAVP 309  
 DB 1653 LFLQVVEL-----QGQVDEQRRREI-----RENTLAEMKRLALISLS 1692  
 QY 310 EDLADR-----VRHGDPVMMVVSQFVKYLIRPQWLEKEIEE 348  
 DB 1693 EDLAHRIEASDKHKKOLEIQALKESSNTELIQNNMA-----LSAMKRVENDEVOI 1743  
 QY 349 ATKKLEKPHIVGNHRTDKVGAEA-----AFHPIEETVHYE-----EDQQLARRMOV- 399  
 DB 1744 ANNELDEYLNELKASEERAKAADADRLAEVRODEHNAVHROKSLLEAKELQAK 1803  
 QY 400 --DKKRYLATDDPALKEAKTKYPSYEFISDSISMSAGLAN--RYTENSILRG 449  
 DB 1804 IDAERAMIOFGAKALAK-----VEDHVSLELMEHSEQRHOSING 1846

RESULT 4  
 SCPL\_MESAU  
 ID SCPL\_MESAU STANDARD: PRT: 845 AA.

AC 060563:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome  
 DE synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome  
 GN SCPL OR SYN.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis; PubMed=7876343;  
 RA MEDLINE=95181577; PubMed=7876343;  
 RA Dobson M.J., Pearlman R.E., Karaliskakis A., Spyropoulos B.,  
 RA Moens P.B.;  
 RT "Synaptonemal complex proteins: occurrence, epitope mapping and  
 RT chromosome disjunction".  
 RL J. Cell Sci. 107:2749-2760(1994).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE (BY SIMILARITY). HAS NON-  
 CC SPECIFIC DNA BINDING CAPABILITY.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR, IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX.  
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC \*\*\*\*\*  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC \*\*\*\*\*  
 DR EMBL: L32978; AAC42039.1; "  
 CC Nucleic protein; Melosis; Cell division; Phosphorylation;  
 KW DNA-binding; Coiled coil.  
 FT NON\_TER 1  
 FT DOMAIN <1 672 COILED COIL (POTENTIAL).  
 FT DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 830 838 ARG/LYS-RICH (BASIC).  
 SQ SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;

Query Match 3.7%; Score 114.5; DB 1; Length 845;  
 Best local similarity 17.6%; Pred. No. 1.9;  
 Matches 78; Conservative 84; Mismatches 181; Indels 99; Gaps 16;

QY 29 LVKNDHSHSSRELSTILAKLERLKOQNDLRMMAGSLRIPGPID--QGASGRVAL 86  
 DB 275 MTEFKNNNEKLEELKILAEQDKLDEKKOYEKLEELQGEQELTLLQTREREVNDL 334  
 QY 87 EDOFMKAKOIEYKQTKNGPKDHETLRRIENGAKELMFLSELKLN----- 139  
 DB 335 EEOULVTKISDQVSKOV-----EELKTLEF-----EKLNAELTASC 373  
 QY 140 ---LEGNEIQRADEFLSLGHNHRSIMT---DLVYLSOTDAGDWREKAKDTELV 191  
 DB 374 GKSLLENKTLQOTNMALAEKKYQEDIDNSKKQBERMKQIENLEE-KETHIRDELEEV 432  
 QY 192 QRRITVLOMPKDC-----SKAKKLVNKN-----GCG-----YGCOL 224  
 DB 433 RKEFIOQGNVEVKCKLDKSEBNARSICEVULKCKOKKILENKCNMLRKOAEKSKYIEEL 492  
 QY 225 HHVYCFMIAVGTOFTLESGINMYATGGMETVFRPV---ETCDRSGSSGTGHSGEV 281  
 DB 493 HQ-----ENKALKKSSAKESKOLNAVLEKVKKIDLELSAKQKQEMTDNQKEI 542  
 QY 282 KQKNQVVELPIVDSVHPRPYLRLAVREDLRLVHGDRAVMMVSQFVKYLIRPQW 341  
 DB 543 EVKKIS--BEKLLGEVEKAKAMWDEAVKLOKEIDELSCQK-----IAMEVNLMEKHNIO 594  
 QY 342 LEKEIEBARKKLGFGKPHIVGNHRTDKVGAEAAPHPIEETVHYNEEDQQLARRMOV-- 399  
 DB 595 YDKIVEERDESLGL-----CKNRDEOQLSVKPA--LETLSNIRNELVLSKOLEIER 645  
 QY 400 -DKKRYLATDDPALKEAKTK 420  
 DB 646 EEKEKLEKEKENTAILKDKDK 667

RESULT 5  
 SH32\_MOUSE  
 ID SH32\_MOUSE STANDARD: PRT: 352 AA.

AC 062420:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE SH3-containing GRB2-like protein 2 (SH3 domain protein 2a) (Endophilin  
 DE 1) (SH3p4).  
 GN SH3GL2 OR SH3D2A.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA MEDLINE=98294438; PubMed=9630982;  
 RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;  
 RT "Cloning of ligand targets: systematic isolation of SH3  
 RT domain-containing proteins".  
 RL Nat. Biotechnol. 14:741-744(1996).  
 CC -1- FUNCTION: PLAYS A ROLE IN SYNAPTIC VESICLE RECYCLING, IN

CC	PARTICULAR IN CLATHRIN-MEDIATED VESICLE ENDOCYTOSIS. EXHIBITS
CC	LYSOPHOSPHATIDIC ACID ACYL TRANSFERASE ACTIVITY (LUPAT) (BY
CC	SIMILARITY)).
CC	-1- SUBUNIT: INTERACTS WITH SYNAPTOTANIN AND DYNAMIN I (BY
CC	SIMILARITY).
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC	-1- SIMILARITY: BELONGS TO THE ENDOPHYLLIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; U58886; AAC71774.1; ..
DR	HSSP; P29355; 1SEM.
DR	MGD; MGI:700009; Sh3d2a.
DR	InterPro; IPR001452; SH3.
DR	Pfam; PF00018; SH3; 1.
DR	PRINTS; PR00452; SH3DOMAIN.
DR	SMART; SM00326; SH3; 1.
DR	PROSITE; PS50002; SH3; 1.
KW	SH3 domain; coiled coil; Multigene family.
FT	DOMAIN 181 248 COILED COIL (POTENTIAL).
FT	DOMAIN 181 248 COILED COIL (POTENTIAL).
FT	DOMAIN 181 248 SH3.
SO	SEQUENCE 352 AA; 39877 MW; EF2ABA34445664B2 CRC64;
Query Match	3.6%; Score 112; DB 1; Length 352;
Blast Local Similarity	19.2%; Pred. No.0.88;
Matches	81; Conservative 60; Mismatches 124; Indels 156; Gaps 18;
OY	178 DWREKAK-DLT-----ELVORRIYIUNPKCKSKAKLVNCI-----NKGCGYGCOL 224
Db	32 DKFKERKVVDYTSRAVMELMTKTEIYLQ-PNPASRAKLSMITGMSKINGKEGPPIY-QA 89
OY	225 HHVVYCEVIAGVTORTLALESNNMRYATGMETVRPVSEFCTSDRSSSTGHMWGEVAKDK 284
Db	90 EALLAEAMLKFG-----RELGGDC-----NEFPALGEVGEA 120
OY	285 NVQAVELTVSVNHRRPYRLPLAVEEDLRVLVRNGSRPAWMVWSQFYKULLIRRPWYLEK 344
Db	121 KREISE--VKDS-----LDMEVKNQNTDRLQNHLKDOL-----R 152
OY	345 ELEEATKKLGEGFNRYIGVNVRTDKVGAEAFNPIEEY-----TVHVED 389
Db	153 ELGNHLKLEGRDLDFGKTKKQKGIIPDEBLQALKEKDESKETALESSNFNLLEDIEGY 212
OY	390 FQLLA---RMNOVDKKRYVALTDDBALAKE-----AKTKYP----SYEFISDNSTS 433
Db	213 SOLSLVAOALEYHNKOAVOILOOVTVRLBERIRQASSQPREYEQKPRMSLEFATGDSYQ .272
OY	434 WSAGLHNRYTESLSGVLIIDHFLSQADFVLTFSQSVCRAVLEYLMQALHPDASANFSL 493
Db	273 PNGSGISHGTIKRP-RGVOMD-----QPCCRALLYD----- 301
OY	494 DDYYEGSPNAHQAIYRNHORPTGEIEMERGDIIGVAGNMWDGYPKGVNKLRGTGLY 553
Db	302 -----EPNEGCELAFKGCDDITTLTNQIDENMYRGMLH--GQSGGF 339
OY	554 P 554
Db	340 P 340
RESULT 6	
RASO_AQDAE	STANDARD:
ID RASO_AQDAE	PRT; 978 AA.
AC 067124;	
DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	

	DE	Probable DNA double-strand break repair rad50 ATPase.	
	CN	RAD50 OR AO_1006.	
	OC	Aquifex aeolicus.	
	CC	Bacteria; Aquificales; Aquificaceae; Aquifex.	
	OX	NCHI_TaxID=63363;	
	[1]		
	RP	SEQUENCE FROM N.A.	
	RC	STRAIN=V95;	
	RA	MEDLINE=98196666; PubMed=9537320;	
	RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,	
	RA	Graham D.E., Overbeek R., Sned M.A., Keller M., Ausley M., Huber R.,	
	RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;	
	RT	"The complete genome of the hyperthermophilic bacterium Aquifex	
	RL	aeolicus."	
	CC	Nature 392:353-358(1998).	
	CC	-I- FUNCTION: Involved in DNA double-strand break repair (DSBR). The	
	CC	rad50/mre11 complex possesses single-strand endonuclease activity	
	CC	and ATP-dependent double-strand-specific exonuclease activity.	
	CC	Rad50 provides an ATP-dependent control of mre11 by unwinding	
	CC	and/or repositioning DNA ends into the mre11 active site (by	
	CC	similarity).	
	CC	-I- SUBUNIT: Forms a complex with mre11 (By similarity).	
	CC	-I- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.	
	CC	-----	
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
	CC	the European Bioinformatics Institute. There are no restrictions on les-	
	CC	use by non-profit institutions as long as its content is in no way	
	CC	modified and this statement is not removed. Usage by and for commercial	
	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
	CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).	
	CC	-----	
	DR	EMBL; AE000718; AAC07092.1; "	
	DR	InterPro: IPR003439; ABC_transport.	
	KW	DNA repair; Hydrolase; ATP-binding; Colloid coll; Complete proteome.	
	FT	NP_BIND 32 39 ATP (BY SIMILARITY).	
	FT	DOMAIN 160 826 COILED COIL (POTENTIAL).	
	SQ	SEQUENCE 978 AA: 115897 MW: 9802BF51ADD1151 CRC64:	
		Query Match 3.6%; Score 111.5; DB 1; Length 978;	
		Best Local Similarity 19.4%; Pred. NO. 3.7;	
		Matches 78; Conservative 74; Mismatches 130; Indels 121; Gaps 18;	
OY	41	RELKTLAKLERLKQCNDLRRMAGSLRIPGPIDGPAISGRALREBOFMKAKEOIEHY 100	
DB	336	KELEHRLKKLQDKEIKELSQLSSLSKEKEREYDA-----KOFEDLSERVEKG 386	
OY	101	KK---QTANGPKDKHEILRRRIENAK---ELMPFLQSEKLKLYLBEGNEIORADEFLS 154	
DB	387	KKLVAETEEKLEKIKELPSEEEYSYLKKAKEKRLVLQRLKLELKEKG-QLENLTQKYKE 445	
OY	155	DLGHRSITMDTVLYLSDTGAGMRKREKAPDTLVQRIRTYLONPPKDCSAKKLYONI 214	
DB	446	KKVKVEK-VLDELKLE-----RELKERLRHNAHMVASYL-SPGD-----TCPV 488	
OY	215	NKSGCGCOLHHVVVCFMIAVGTOFTLLAESHNMYATGCMETVERPVSETCTDRSGSST 274	
DB	489	---CG-----GIYGKALEN-----VDAGEIGSE 508	
OY	275	GHWSEGVCKNOVVLDPI-----VDSVHPRPYPLAVPEIDLADR----- 316	
DB	509	LKHAELEKEBEREDITDTLKLYAQKINSLEKEMEKRLAREVELREIKEIDPENLKERIKKLEE 568	
OY	317	VRVHDDPAAVMVVSQPVYKTLIRPOPLKEIKIEIATKGLGFKNHVGVNHRT-----DKV 370	
DB	569	LRIEKEKLEHKLKRYRKALEDRO----KKEEAOKL-----HKAOTELELLKERT 615	
OY	371	GAEAAF-----HPIEETYVHVEDFOFLARRMO-VDRK 402	
DB	616	REKSRLVKEFKELYRVERLEDYEESLKKEINININSKLOEIEEK 658	

CARB\_BACSU  
ID CARB\_BACSU STANDARD: PRT: 1071 AA.  
AC P25994:  
DT 01-MAY-1992 (rel. 22, Created)  
DT 01-MAY-1992 (rel. 22, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Carbamoyl-phosphate synthase, pyrimidine-specific, large chain  
DE (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).  
GN PYRAB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_Taxid=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91225016: PubMed-1709162;  
RX Quinn C.L., Stephenson B.T., Switzer R.L.;  
RT "Functional organization and nucleotide sequence of the Bacillus  
RT subtilis pyrimidine biosynthetic operon.";  
RT J. Biol. Chem. 266:9113-9127(1991).  
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
CC phosphate + glutamate + carbamoyl phosphate.  
CC -1- PATHWAY: INVOLVED IN PYRIMIDINE BIOSYNTHESIS.  
CC -1- SUBUNIT: COMPOSED OF TWO CHAINS; THE SMALL (OR GLUTAMINE) CHAIN  
CC PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED  
CC BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M59757; AAA21270.1; -  
DR EMBL: 299112; CAB13426.1; -  
DR PIR: F39845; F39845.  
DR HSSP: F00968; IAX9.  
DR Subtilist: BG10716; PYRAB.  
DR Interpro: IPR000384; MG5.  
DR Interpro: IPR000384; MG5.  
DR Pfam: PF00289; Cpsase\_L\_chain: 2.  
DR Pfam: PF02786; Cpsase\_L\_D2: 2.  
DR Pfam: PF02787; Cpsase\_L\_D3: 1.  
DR Pfam: PF02142; MG5: 1.  
DR PRINTS: PR00098; CPSASE.  
DR PROSITE: PS00866; Cpsase\_1; 2.  
DR PROSITE: PS00867; Cpsase\_2; 2.  
KW Pyrimidine biosynthesis; Ligase; Repeat; ATP-binding;  
KW Complete proteome.  
FT REPEAT 1 544  
FT REPEAT 545 1071  
FT NP\_BIND 153 210 ATP (POTENTIAL).  
FT NP\_BIND 302 352 ATP (POTENTIAL).  
FT SEQUENCE 1071 AA; 117649 MW; 195E33CFC562222C CRC64;  
Query Match 3.6%; Score 111.5; DB 1; Length 1071;  
Best Local Similarity 19.0%; Pred. No. 4.2; Indels 219; Gaps 29;  
Matches 119; Conservative 83; Mismatches 204;  
43 LSKLAKLEKLEKQONEDLEKRMAGSL-----RIPEGPID-----OGP 78  
DB 318 IAKLAKIAVGLSLDEMMNPVTKTYAFAPEPALDYVVSIIIPMPFKFESANRKLGTQAK 377  
QY 79 ASGRV-----RALEOPFMKKEQIEN--YKQTKNGPGKGEHILRRIRIENGAKELMFLOS 132  
DB 378 AGEVVAIRTEELSLKAVRSLEADVYHLELEKADADISDELEKRIKACDEERLEFLAE 437  
QY 133 ELKKLKNLEGNELQRIHADFLSDLGHERSIMDLYLISOTGS-----AGD----- 178  
DB 438 AVRGGTVE-----DL--HEPSAI-DVEFLKLFGLVIOFEKELKANGDTDV 481

QY 179 -----WREKAKDLTELVORRI-----TYL 198  
DB 482 LRAKELGFSDDYISREMKKESLYSLRKQAGIAFVKFMVDTCAEFESSEPTFYSTYE 541  
QY 199 QNPKDCSKAKKLYCINIKG---CGYGCQIHHVYVCPMAIYQGRFLALESNNMRKATGCM 255  
DB 542 EENESVYTDKKSVMLVSGSPIRIGGVEDP-----YATVHSV-----MAIKQAGY 586  
QY 256 ETVF--RPVSECTPDRSGSGSTGHWGGEKKNVQVELPIYVSVPRPPYPLAV----- 308  
DB 587 EAIIVNNNEPTVSTDS-----ISDK-LYEPRLTEDEVMIIIDLEGPMGVYVQFG 635  
QY 309 -----PEDLADRLVRVHGDPVAVMWVSQFYLLIRPQWMEKEIEATKLGFKHPV-- 359  
DB 636 GGTAINLADDELARGVKILGTS-----LEDDLRABD--RDKFEQALGELGVQRPGLK 685  
QY 360 TGVHVRIRDKVGAEAFAHPIEETVHVEDFOLLARRQOVCKRRYYLATDPAALLKEAKT 419  
DB 686 TATSVNQAVSIAISDIG-----YPLVLRPSYVLGGRAMEI---YHHEBELHMYKNAVK 735  
QY 420 KPSVEEFDISNLSMSAGLHNRVTENSLRGVILDIHFLSQADEFLCTESSQVCRAVEIM 479  
DB 736 IMPQHPVLD-----RY-----LTGKEIEVDVSDGETVYIP-----GIM 770  
QY 480 QALHPDASANFRSLDDIYVFGSPNANHQIATVHPQRTG-----EIPMEPG-DIIG 530  
DB 771 EHIER-----ACVHSQDSIAVYRPSLTFEDIKKIEQYITLAKGLNIYG 815  
QY 531 VAG-----NHMDGKRGVNRKLGRT 550  
DB 816 LNIQFVLSQGEVYVLEVPNRSSRT 840  
RESULT 8  
ID PDL1\_HUMAN STANDARD: PRT: 1074 AA.  
AC Q13593:  
DT 01-NOV-1997 (rel. 35, Created)  
DT 01-NOV-1997 (rel. 35, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Phospholipase D1 (EC 3.1.4.4) (PLD 1) (choline phosphatase 1)  
DE (Phosphatidylcholine-hydrolyzing phospholipase D1) (mPLD1).  
GN PDL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM PDL1A).  
RX MEDLINE-96102003: PubMed-8530346;  
RA Hammond S.M., Altschuller Y.M., Sung T.-C., Rudge S.A., Rose K.,  
RA Engelbrecht J., Morris A.J., Frohman M.A.;  
RT "Human Adp-ribosylation factor-activated phosphatidylcholine-specific  
RT phospholipase D defines a new and highly conserved gene family.";  
RT J. Biol. Chem. 270:29640-29643(1995).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS PDL1A AND PDL1B).  
RX MEDLINE-97166247: PubMed-9013646;  
RA Hammond S.M., Jenco J.M., Nakashima S., Cadwallader K., Gu Q.-M.,  
RA Cook S., Nozawa Y., Prestwich G.D., Frohman M.A., Morris A.J.;  
RT "Characterization of two alternately spliced forms of phospholipase  
RT D1. Activation of the purified enzymes by phosphatidylinositol 4,5-  
RT biphosphate, Adp-ribosylation factor, and Rho family monomeric GTP-  
RT binding proteins and protein kinase C-alpha.";  
RT J. Biol. Chem. 272:3860-3868(1997).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS PDL1A; PDL1B AND PDL1C).  
RX TISSUE=Cervical carcinoma, Skeletal muscle, Chondrocytes, and Brain;  
RX MEDLINE-98437320: PubMed-9761774;  
RA Steed P.M., Clark K.L., Boyar W.C., Lasala D.J.;  
RT "Characterization of human PLD2 and the analysis of PLD isoform splice  
RT variants.";  
RT FASEB J. 12:1309-1317(1998).

[4]  
RP SEQUENCE OF 739-1074 FROM N.A. (ISOFORM PLD1D).  
RA Hughes W.E., Parker P.J.:  
RT "A novel human phospholipase D1 splice variant displays conserved  
RL regulation in vitro but altered localisation in vivo."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=98250727; PubMed=9582313.  
RA Lopez I., Arnold R.S., Lambeth J.D.:  
RT "Cloning and initial characterization of a human phospholipase D2  
RL (hPLD2). ADP-ribosylation factor regulates hPLD2."  
RL J. Biol. Chem. 273:12846-12852(1998).  
CC -1- FUNCTION: IMPLICATED AS A CRITICAL STEP IN NUMEROUS CELLULAR  
CC PATHWAYS, INCLUDING SIGNAL TRANSDUCTION, MEMBRANE TRAFFICKING, AND  
CC THE REGULATION OF MITOSIS. MAY BE INVOLVED IN THE REGULATION OF  
CC PERINUCLEAR INTRACELLULAR MEMBRANE TRAFFIC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
CC phosphate.  
CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-  
CC BISPHOSPHATE AND PHOSPHATIDYLINOSITOL 3,4,5-TRISPHOSPHATE,  
CC ACTIVATED BY THE PHOSPHOKINASE C-ALPHA, BY THE ADP-RIBOSYLATION  
CC FACTOR-1 (ARF-1), AND IN A LESSER EXTENT BY GTP-BINDING PROTEINS:  
CC RHO A, RAC-1 AND CDC42. INHIBITED BY OLEATE.  
CC -1- SUBCELLULAR LOCATION: PERINUCLEAR REGIONS; ENDOPLASMIC RETICULUM,  
CC GOLGI APPARATUS AND LATE ENDOSOMES (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: PLD1A (SHOWN HERE), PLD1B, PLD1C  
CC AND PLD1D; ARE PRODUCED BY ALTERNATIVE SPLICING. PLD1C IS UNLIKELY  
CC TO BE PHYSIOLOGICALLY RELEVANT.  
CC -1- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY IN THE PANCREAS AND  
CC HEART AND AT HIGH LEVELS IN BRAIN, PLACENTA, SPLEEN, UTERUS, AND  
CC SMALL INTESTINE.  
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 PHO HOMOLOG (PX) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 PLDC DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: U38545; AAB49031.1; .  
CC EMBL: AJ276230; CAB76564.1; .  
CC MIM: 602382; .  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR001736; PLD.  
DR InterPro: IPR001683; PX.  
DR Pfam: PF00169; PH; 1.  
DR Pfam: PF00614; PLDC; 2.  
DR Pfam: PF00787; PX; 1.  
DR SMART: SM00233; PH; 1.  
DR SMART: SM00155; PLDC; 2.  
DR SMART: SM00312; PX; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; FALSE\_NEG.  
KW Hydrolase; Lipid degradation; Membrane; Alternative splicing;  
KW Endoplasmic reticulum; Golgi stack; Repeat.  
FT DOMAIN 76 196  
FT DOMAIN 219 328  
FT DOMAIN 459 486  
FT DOMAIN 891 918  
FT DOMAIN 928  
FT VARSPLIC 514 597  
FT  
FT  
FT  
FT  
FT VARSPLIC 598 1074  
FT VARSPLIC 585 623

FT VARSPLIC 962 971 VILGYLDPS -> SKMTPEVDEP (IN ISOFORM  
FT PLD1D).  
FT VARSPLIC 972 1074 MISSING (IN ISOFORM PLD1D).  
FT CONFLICT 832 832 S -> P (IN REF. 3).  
SQ SEQUENCE 1074 AA: 124184 MW: 734F285790ADBFE7A CRC64:  
Query Match 3.6%: Score 110.5; DB 1; Length 1074;  
Best Local Similarity 18.0%: Pred. No. 4.9;  
Matches 101; Conservative 67; Mismatches 166; Indels 227; Gaps 24:  
QY 114 ILRRRIENGAK-----ELMFFLOSELKK--LKNLEGN-ELQRRADEFLSDL---GH 158  
D 114 ILRRRIENGAK-----ELMFFLOSELKK--LKNLEGN-ELQRRADEFLSDL---GH 158  
D 404 ILKRAQGVNIFIMLYKEVELLAGINSEYTKRLMRHPKIKVMRHPDHVSSVYIMAH 463  
QY 159 HERSIMTD--LYVISQDGA-GDMREKAKDLTELVORRIYVIONPKDCSAKRLVCNIN 215  
D 159 HERSIMTD--LYVISQDGA-GDMREKAKDLTELVORRIYVIONPKDCSAKRLVCNIN 215  
D 464 HEKLIIDQSAFVFGIDILAYGRMDNE-----HRLDVGSVVRKVTGSPSL----- 509  
QY 216 KGCYGCGLHVVYCFMFAVGTOFTLALESHNMRYATGGEVFRPVSECTDSSSGSTG 275  
D 216 KGCYGCGLHVVYCFMFAVGTOFTLALESHNMRYATGGEVFRPVSECTDSSSGSTG 275  
D 510 -----GSLPPAAMESME----- 521  
QY 276 HMGSEVKDKNVVVELPI-----VDS----- 296  
D 276 HMGSEVKDKNVVVELPI-----VDS----- 296  
D 522 --SLRLKDKNEPVONLPIQKSIDVDLSKLGIGKPRKFSKSLYKOLRRHLLHDADSISS 579  
QY 297 -----VHPPRYLPPLAVPEDLADR-LVR-----VHG 321  
D 297 -----VHPPRYLPPLAVPEDLADR-LVR-----VHG 321  
QY 580 IDSVSYENHYRSHNHLHGKLPKLFHPSESEOGTLRPHADTGSIRSLQTVGELHG 639  
D 580 IDSVSYENHYRSHNHLHGKLPKLFHPSESEOGTLRPHADTGSIRSLQTVGELHG 639  
QY 322 DPVAVWWSQFVKYLLIRPQMLEKELEE-----ATKKLGFKRPVGVNHRRTDKGAELAF 376  
D 322 DPVAVWWSQFVKYLLIRPQMLEKELEE-----ATKKLGFKRPVGVNHRRTDKGAELAF 376  
D 640 ETRFMHGKDYCNFVKDWQVLDKPRADFDIRYSTRPMWHDIASAVH---GKAARDVAR 695  
QY 377 HPLEETVHVEEDYQOLARRQVQDKRYVLTADDPALKEAKTKYSPSEFISDSI----- 432  
D 377 HPLEETVHVEEDYQOLARRQVQDKRYVLTADDPALKEAKTKYSPSEFISDSI----- 432  
D 696 HPIQRW-----NFTKI-----MKSKYRSLSYRPLPKSQTANHELRYQVPSYAHNV 742  
QY 433 -----SWAGCLNKRRTENSRLGVLDI-----HFLSQADFLVCFPSOVC 472  
D 433 -----SWAGCLNKRRTENSRLGVLDI-----HFLSQADFLVCFPSOVC 472  
D 743 QLLRSAAWMSAGI--KYEEESIHAAVYHVIENSRYIYIENQFISCADKRV--YFNKIGD 799  
QY 473 RVAREIMQA-----LHPDASANFRSLDIIYGGNANNAHQIAYHPORTE 518  
D 473 RVAREIMQA-----LHPDASANFRSLDIIYGGNANNAHQIAYHPORTE 518  
D 800 AIDQRIILAHRENOKRYRVYVYIPLPGEGDIST-----GGNALQAIHMFYRTWCR 852  
QY 519 GEIPMEPGDIIGVAGNHWDGY 539  
D 519 GEIPMEPGDIIGVAGNHWDGY 539  
D 853 GENST-LGQLKAEELGNQWINT 872  
RESULT 9  
SCPL\_HUMAN STANDARD; PRT; 976 AA.  
ID SCPL\_HUMAN 015431: 014963:  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Synaptonemal complex protein 1 (SCP-1 protein).  
GN SCPL OR SCPL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=97224467; PubMed=9119375;  
RA Meuwissen R.L.J., Weerts I., Hoovers J.M.N., Lescot N.J.,  
RT Heyting C.:  
RT "Human synaptonemal complex protein 1 (SCP1): Isolation and  
RT characterization of the cDNA and chromosomal localization of the  
RT gene.";

RL Genomics 39:377-384(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Testis;  
 RX MEDLINE=98037449; PubMed=9371398;  
 RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsunaga K., Namiki M.,  
 RA Okuyama A., Nishimune Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCP1) to human  
 RT chromosome 1p13 by fluorescence in situ hybridization and its  
 RT expression in the testis";  
 RL Cytogenet. Cell Genet. 78:103-104(1997).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 DR EMBL: X95654; CAA64956.1; -  
 DR EMBL: D67035; BAA22586.1; -  
 DR MIM: 602162; -  
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 KM DNA-binding; Coiled coil.  
 FT DOMAIN 107 120  
 FT 107 798  
 FT 117 120  
 FT 679 682  
 FT 880 883  
 FT 961 969  
 FT DOMAIN 46 46  
 FT 106 106  
 FT 153 153  
 FT 161 161  
 FT 168 168  
 FT 216 216  
 FT 225 226  
 FT 350 350  
 FT 360 360  
 FT 400 400  
 FT 406 406  
 FT 415 415  
 FT 449 449  
 FT 483 510  
 FT 516 528  
 FT 549 549  
 FT 560 560  
 FT 805 805  
 FT 941 941  
 SQ SEQUENCE 976 AA: 114069 MW: 88A8ID042AC2696B CRC64;  
 Query Match 3.6%; Score 110; DB 1; Length 976;  
 Best local Similarity 20.5%; Pred. No. 4.7;  
 Matches 83; Conservative 69; Mismatches 152; Indels 100; Gaps 16;  
 40 SRRLSLAKLRLKQONEDLRHAGSLRIPGIDGPAAGRVRLAEQFMKAKBOIEN 99

Db 496 SKVEKDLTELENEKLNKLTSHCNKLSLENKELTQ-ETSDMTLELKNQ----QEDINN 550  
 Qy 100 YKKQTKNGCKKHHELLRRIENGAKELMFWLOSCLKKLNKNGNELQRIADFEISDLGHH 159  
 Db 551 NKKQ-----EENRLKQIEN-LOETETQRLNELEYR---EELKKQKRDCKKL- 594  
 Qy 160 ERSIMTDLXYLSQTDGAGCMRKEAKDLTELVRRTTYLQNPDKCSKAKKLCVINKCGC 219  
 Db 595 -----DKSEENCNNLKQYENKKKYIEELQGENKALK-----KKQTA 631  
 Qy 220 YCCQLHHVYCFMAYGTOFTALESHNRVATGNETVFRPVSETCTDRSSSTGHSNG 279  
 Db 632 ESKQLN--VEYELKV--NKLELELESARKQF-----GEITDTYOK 666  
 Qy 280 EYKDKVQVVELPIYDSVHPRPRLPLANPEDLDLVRVHGD---PAVWWSQFVKYLI 336  
 Db 667 ELEDKRIIS--EENLLEVEK-----AKVIADAVVLQKEIDRCQHKIAENVALME 715  
 Qy 337 RPOWLEKEIEEATKKGKHNPIGVHVRPTDVGAEAFHPIEETVAVEEDFOLLAR 396  
 Db 716 KKHQYDKIIEERDSEL-----GLYKSEDOSSLRSLELSNLKAE-----LLSVK 764  
 Qy 397 MOVD-----KKRYLATDPPALLKEAKTKYPSYEISONSISW 434  
 Db 765 KOLEIEREKEKLRKREKENTATLKEKKDKKQTQTELETPEIYW 808  
 RESULT 10  
 SCPL\_MOUSE  
 ID SCPL\_MOUSE STANDARD; PRT: 993 AA.  
 AC Q62209; Q62329; O09205; P70192;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SCPL OR SCPL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA; TISSUE=Testis;  
 RX MEDLINE=96004899; PubMed=7548215;  
 RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;  
 RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";  
 RL Biochim. Biophys. Acta 1263:258-260(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS; TISSUE=Testis;  
 RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;  
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-149 FROM N.A.  
 RC STRAIN=C57BL/6;  
 RA Hoge J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,  
 RA Sage C., Cuzin F., Rassoulzadegan M.;  
 RL Submitted (Jan-1997) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 95-787 FROM N.A.  
 RC STRAIN=ICR; TISSUE=Testis;  
 RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;  
 RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).



CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: 238118; CAA86262.1; -  
DR EMBL: U41069; AAA64514.1; ALT\_INTF.  
DR EMBL: U62864; AAC53335.1; -  
DR EMBL: U62860; AAC53335.1; JOINED.  
DR EMBL: U62861; AAC53335.1; JOINED.  
DR EMBL: U62862; AAC53335.1; JOINED.  
DR EMBL: U62863; AAC53335.1; JOINED.  
DR EMBL: D88539; BAA13639.1; -  
DR MGI: 105931; Sycepl.  
KW Nuclear protein; Melosis; Cell division; Phosphorylation;  
KW DNA-binding; Coiled coil.  
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 104 815 COILED COIL (POTENTIAL).  
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT CONFLICT 527 527 F -> L (IN REF. 2).  
SQ SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAFE6 CRC64;

Query Match 3.6%; Score 110; DB 1; Length 993;  
Best Local Similarity 18.9%; Pred. No. 4.8;  
Matches 85; Conservative 78; Mismatches 175; Indels 112; Gaps 17;

QY 42 ELSTLAKLERKQONEDRRMAGSLRIPECPID--QGPAQRVRLAEQFKAKKEQIEN 99  
DB 432 ELKNIILAEQKILDEKQYKLAELQEKQELTFLREKREKVIDLQYVYVTKTSEH 491  
QY 100 YKQTRNGPGKDHETLRRRIENGAKELMFLQSELKILN-----LEGNELORH 148  
DB 492 YLKQ-----VEEMKTEL-----EKELKNTELTASCDMLLENKKFVQE 530  
QY 149 ADEFLSDGCHHRSTMT---DLVYISQDAGDMREKAKDLTELVORRTYLLQNPKD- 203  
DB 531 ASDMALTELKHOEDINCKKOEKRLKQIE---NLEEKEMHLRDELSVRKEFIQGGDEV 587  
QY 204 -----CSKAKKILVCNINKGGYGCOLHHVYCFM 232  
DB 588 KKKLKSEENARSICEVUKKEKOKILKSCNNLKKOVENSKNIE---ELHQ----- 638  
QY 233 IAYGTQRTLALESNMRVATGMEVTFRPS---ETCTDRSSSTGSHMSGVKDKNVQV 289  
DB 639 -----ENKTLTKKSSAEIKOLNAYEIKVSKLELELESTKQRFEMTNVNYQKEIENKIS-- 692  
QY 290 ELPIYDSVHPRPYPLAIPEDLADRLVAVHGDPAVMVVSQVVKILIRPQMLEKITEEA 349  
DB 693 ESKLLGEVEKAKATYDEAKIQLQEKIDLRQKH-----IAENVALMEKKHKOYDKIVEER 746  
QY 350 TKKLG-FKHPIVGVHVRITDKVGAEAAPHIEEYVHVEDEDQLARRQV---DKKRVY 405  
DB 747 DSELGLIKN-----REEQSSAKIA---LETLSININELVSLKKQLEKEKEKELK 796  
QY 406 LATDDPALIKEAKTYRPSYEFT-SDNSISW 434  
DB 797 MAKENTAILDKDKKKIKQASLSLESPATSW 826

RESULT 11  
CNF\_HUMAN STANDARD; PRT; 3210 AA.  
ID CNF\_HUMAN  
AC P49454; Q13246; Q13171;

DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE CENP-F kinetochore protein (Centromere protein F) (Mitosin) (AH  
DE antigen).  
GN CENP-F.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Breast carcinoma;  
RX MEDLINE=95348175; PubMed=7542657;  
RA Liao H., Winkfeld R.J., Mack G., Rattner J.B., Yen T.J.;  
RT "CENP-F is a protein of the nuclear matrix that assembles onto  
RT kinetochores at late G2 and is rapidly degraded after mitosis";  
RL J. Cell Biol. 130:507-518(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95379848; PubMed=7651420;  
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,  
RA Jones D., Yang-Feng T.L., Lee W.-H.;  
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein  
RT that is specifically involved in mitotic-phase progression";  
RL Mol. Cell. Biol. 15:5017-5029(1995).  
RN [3]  
RP SEQUENCE OF 2194-3210 FROM N.A.  
RX MEDLINE=95336446; PubMed=7612011;  
RA Li Q., Ke Y., Kapp J.A., Partig N., Medsker T.A. Jr., Joshi H.C.;  
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal  
RT domain sufficient for nuclear localization";  
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=95370296; PubMed=7642639;  
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;  
RT "The C terminus of mitosis is essential for its nuclear localization,  
RT centromere/kinetochore targeting, and dimerization";  
RL J. Biol. Chem. 270:19545-19550(1995).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=98437347; PubMed=9763420;  
RA Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
RT interactions with the kinetochore proteins CENP-F and hBUBR1";  
RL J. Cell Biol. 143:45-63(1998).  
CC -1- FUNCTION: PROBABLY REQUIRED FOR KINETOCORE FUNCTION, INVOLVED IN  
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH  
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.  
CC -1- SUBUNIT: HOMO- OR HETERODIMER.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),  
CC REORGANIZATION TO THE KINETOCORE/CENTROMERE (CORONAL SURFACE OF  
CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.  
CC -1- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.  
CC -1- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U19769; AAA82889.1; -  
DR EMBL: U30872; AAA82935.1; -  
DR EMBL: U25725; AAA86839.1; -  
DR HSP: P02649; ILE4.  
DR MIM: 600236; -  
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;  
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.  
FT DOMAIN 14 197 COILED COIL (POTENTIAL).



DB 230 -----LTRLIR-----EVEKILLANGFEKVKSP 252  
OY 359 VIGVHVRTDKGAEAFHFEYTVHVEDPQLLARWQVKKRYLATDPPALKEAK 418  
DB 253 IVETEFMNFDM-L-FEPQHF-----AREMQ-----DTPFL-----281  
OY 419 TTPSYEFISDINSISASQILNHR-----YTENSLRGVILIDIFLSQLADEFLVCTFSSQVCR 473  
DB 282 -KYPNEGDIPELLLSKVKEVHERCKMYKDENVSRLLRTHTTASSINRYLASLSDEEKN 340  
OY 474 VAYET--MOALHPDASANFESLDDIYFEGC-----PNAHNOI-----508  
DB 341 KEHVKVFCIDRVERNEAIDKHLPEFYOCGCIIMDDNVNENLIGVLKEFLNRLGFEKVAF 400  
OY 509 --AIVPH-OPRTGELIPMEFGDILIGVAGNHM 536  
DB 401 RPAVFPTEPESLDAEYVE-----GKGW 423

RESULT 13  
OSFL\_HUMAN  
ID OSFL\_HUMAN STANDARD: PRT: 214 AA.  
AC 092882;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Osteoclast stimulating factor 1.  
CN OSTF1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Bone marrow;  
RA MEDLINE=99190637; PubMed=10092216;  
RA Reddy S.V., Devlin R., Mena C., Nishimura R., Choi S.J., Dallas M.,  
RA Yoneda T., Roodman G.D.;  
RT "Isolation and characterization of a cDNA clone encoding a novel  
peptide (OSF) that enhances osteoclast formation and bone  
resorption.";  
RT J. Cell. Physiol. 177:636-645(1998).  
RN [2]  
RP SMN1 BINDING.  
RA Roodman G.D.;  
RL Unpublished observations (NOV-2000).  
CC -1- FUNCTION: INDUCES BONE RESORPTION, ACTING PROBABLY THROUGH A  
CC SIGNALING CASCADE WHICH RESULTS IN THE SECRETION OF FACTOR(S)  
CC ENHANCING OSTEOCLAST FORMATION AND ACTIVITY.  
CC -1- SUBUNIT: BINDS C-SRC, AND SMN1.  
CC -1- TISSUE SPECIFICITY: URICITOUSLY EXPRESSED.  
CC -1- SIMILARITY: CONTAINS 3 SH3 REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U63717; AAB06396.1; -  
DR HSP: P06241; ISHP.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF000023; ank; 3.  
DR Pfam: PF00018; SH3; 1.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR SMART: SM00248; ANK; 2.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS50002; SH3; 1.

DR PROSITE: PS50297; ANK\_REPEAT; 1.  
DR PROSITE: PS50088; ANK\_REPEAT; 1.  
KW SH3 domain; Repeat; ANK repeat.  
FT DOMAIN 4 11 PRO-RICH.  
FT DOMAIN 12 71 SH3.  
FT REPEAT 72 101 ANK 1.  
FT REPEAT 105 135 ANK 2.  
FT REPEAT 139 168 ANK 3.  
SQ SEQUENCE 214 AA: 23799 MW: E10C6570B68E3D6D CRC64;  
Query Match 3.5%; Score 109; DB 1; Length 214;  
Best Local Similarity 38.0%; Pred. No. 0.74;  
Matches 27; Conservative 9; Mismatches 25; Indels 10; Gaps 2;  
OY 509 AIVPHOPRTGELIPMEFGDILIGVAG-----NHWDTGPKGVNRLGTGLTPSKVKEKLET 564  
DB 19 ALVTEPRTPELDYFEEGDIIYITDMSDTNMWKGTSK-----GRTGLIPSYVAEQAES 72  
OY 565 VKPTYPEADK 575  
DB 73 IDNPLHEAKR 83

RESULT 14  
NODZ\_BRAJA  
ID NODZ\_BRAJA STANDARD: PRT: 324 AA.  
AC 045271;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Nodulation protein 2 (EC 2.4.1.-).  
GN NODZ.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Bradyrhizobium group; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RA MEDLINE=94131940; PubMed=8300517;  
RA Stacey G., Luka S., Sanjuan J., Bantlavi Z., Nieuwkoop A.J.,  
RA Chun J.Y., Forsberg L.S., Carlson R.;  
RT "nodZ, a unique host-specific nodulation gene, is involved in the  
RT fucosylation of the lipooligosaccharide nodulation signal of  
RT Bradyrhizobium japonicum.";  
RL J. Bacteriol. 176:620-633(1994).  
CC -1- FUNCTION: FUCOSYLTRANSFERASE WHICH ADDS THE FUCOSE MOIETY OF THE  
CC NOD FACTOR ON ITS TERMINAL REDUCING N-ACETYLGLUCOSAMINE END. USES  
CC GDP-FUCOSE AS THE DONOR GROUP.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L22756; AAB58813.1; -  
KW Nodulation; Transferase; Glycosyltransferase.  
SQ SEQUENCE 324 AA: 36668 MW: 0DA83CFB9E040E22 CRC64;  
Query Match 3.5%; Score 109; DB 1; Length 324;  
Best Local Similarity 21.2%; Pred. No. 1.3; Mismatches 10; Gaps 14;  
Matches 68; Conservative 48; Mismatches 10; Indels 94; Gaps 14;  
OY 182 KEAKDLTELQRRITLYIQPKDSKAKKLVNCIN-----KGGCGCOLHHVVCFWI 233  
DB 23 RETSVLTSLVQ-----FGAREKARQVSSGSSNDPRFVSSRRRGFGGCLSLAAAMF 74  
OY 234 AVGTORTALESNMRYA-----TGGMETVFRPVSE-----TCTDRSGSSTGHWGCE 280  
DB 75 AKOTGRTLAID--WRGSCYLDPEPTNAPVFPFEVEDIGGVAVICDD-----119





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 : Search time 22 seconds  
(without alignments)  
2511.426 Million cell updates/sec

Title: US-09-839-136-2

Perfect score: 3096

Sequence: 1 MRPWGSMRWIMLILFAWGT.....YKVKRIETVKKYPPPEADK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2949	95.3	575	1 JCS432	glycoprotein 6- $\alpha$ p
2	948	30.6	818	2 T32154	hypothetical prote
3	140	4.5	1025	2 H86250	hypothetical prote
4	123.5	4.0	1057	2 F89892	cardamoyl-phosphat
5	123	4.0	1871	2 D96796	probable heat shoc
6	122.5	4.0	1427	2 S22695	reslin - human
7	121.5	3.9	1154	2 F71856	hypothetical prote
8	118.5	3.8	1392	2 M43336	microtubule-vesicl
9	117.5	3.8	2094	2 S33124	lpt protein - huma
10	117	3.8	1938	1 MKWK1	myosin heavy chain
11	116	3.7	544	2 T32568	hypothetical prote
12	114.5	3.7	845	2 I48176	synaptonemal compl
13	112	3.6	978	2 H89971	conserved hypotnet
14	111.5	3.6	978	2 A70387	conserved hypotnet
15	111.5	3.6	1071	2 F39845	cardamoyl-phosphat
16	111	3.6	398	2 T04501	hypothetical prote
17	111	3.6	570	2 C90485	beta-glucuronidase
18	111	3.6	1011	2 T13055	dynamn associated
19	111	3.6	1094	2 T13053	dynamn associated
20	111	3.6	1549	2 T21809	hypothetical prote
21	111	3.6	1940	2 A59287	myosin heavy chain
22	110.5	3.6	909	2 T00009	probable primase (
23	110	3.6	993	2 S49461	synaptonemal compl
24	109.5	3.5	480	2 G64360	phenylalanine--L-RN
25	109.5	3.5	1179	2 F71190	probable chromosom
26	109	3.5	324	2 A36959	host-specific nodu
27	109	3.5	417	2 D75183	hypothetical prote
28	109	3.5	822	2 G69010	conserved hypotnet
29	108.5	3.5	1510	2 T16927	hypothetical prote

30	108.5	3.5	1738	2 T14867	interaplin - slime
31	108	3.5	678	2 H88187	protein C18H9.8 (1
32	107.5	3.5	503	4 S23741	hypothetical TRP/T
33	107.5	3.5	2288	2 T29999	hypothetical prote
34	107	3.5	2441	2 D71623	erythrocyte membra
35	106.5	3.4	900	2 T19689	hypothetical prote
36	106.5	3.4	2326	2 T29140	hypothetical prote
37	106	3.4	581	2 B87768	protein P54C1.1 (1
38	106	3.4	1992	1 S02771	myosin heavy chain
39	105.5	3.4	487	1 T32635	phosphoprotein pho
40	105.5	3.4	646	2 A44985	dnak-type molecula
41	105.5	3.4	758	2 H87133	methylnalonyl-CoA
42	105.5	3.4	1188	2 E89896	chromosome segrega
43	105	3.4	674	1 DAHUAL	archidonate 5-11p
44	105	3.4	880	2 F75103	conserved hypotnet
45	105	3.4	919	2 B72765	hypothetical prote

ALIGNMENTS

RESULT 1

JCS432

glycoprotein 6- $\alpha$ -L-fucosyltransferase (EC 2.4.1.68) precursor - human

N:Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6- $\alpha$ -

C:Species: Homo sapiens (man)

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000

C:Accession: JCS432; PC4322

R:Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.

J. Biochem. 121, 626-632, 1997

A:Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alp

A:Reference number: JCS432; MUID:97279058

A:Accession: JCS432

A:Molecule type: mRNA

A:Residues: 1-575 <YAN1>

A:Cross-references: DDBJ: D89289; NID: g2055306; PID: BAA19764.1; PID: g2055307

A:Accession: PC4322

A:Molecule type: protein

A:Residues: 68-87:352-376:419-432 <YAN2>

C:Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to as

C:Gene: GDB: FUT8

A:Cross-references: GDB: 9786294; OMIM: 602589

A:Map position: 14q23-14q23

C:Superfamily: human glycoprotein 6- $\alpha$ -L-fucosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-575/Product: glycoprotein 6- $\alpha$ -L-fucosyltransferase #status predicted <MAT>

Query Match 95.3% Score 2949; DB 1; Length 575;

Best Local Similarity 95.5% Pred. No. 1.3e+197;

Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY	1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSRELSKTLAKLEKQONEDL 60		1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSRELSKTLAKLEKQONEDL 60
DB	1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSRELSKTLAKLEKQONEDL 60		1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSRELSKTLAKLEKQONEDL 60
QY	61	RMASGLRIPEDIPDGPASGVRALEQFMKAEQIENYKKQTKNPGCKDHEILRRRIE 120		61	RMASGLRIPEDIPDGPASGVRALEQFMKAEQIENYKKQTKNPGCKDHEILRRRIE 120
DB	61	RMASGLRIPEDIPDGPASGVRALEQFMKAEQIENYKKQTKNPGCKDHEILRRRIE 120		61	RMASGLRIPEDIPDGPASGVRALEQFMKAEQIENYKKQTKNPGCKDHEILRRRIE 120
QY	121	NAKELMPFLQSELEKLLKLNLEGNELORHNADEFLSDIGHNERSIMPDLYLSTQDGDGR 180		121	NAKELMPFLQSELEKLLKLNLEGNELORHNADEFLSDIGHNERSIMPDLYLSTQDGDGR 180
DB	121	NAKELMPFLQSELEKLLKLNLEGNELORHNADEFLSDIGHNERSIMPDLYLSTQDGDGR 180		121	NAKELMPFLQSELEKLLKLNLEGNELORHNADEFLSDIGHNERSIMPDLYLSTQDGDGR 180
QY	181	EKEADLTELVRITTYLQNPDKCSAKKLYCNINKGCYGCQLHHVVYCFMIAVGTQRT 240		181	EKEADLTELVRITTYLQNPDKCSAKKLYCNINKGCYGCQLHHVVYCFMIAVGTQRT 240
DB	181	EKEADLTELVRITTYLQNPDKCSAKKLYCNINKGCYGCQLHHVVYCFMIAVGTQRT 240		181	EKEADLTELVRITTYLQNPDKCSAKKLYCNINKGCYGCQLHHVVYCFMIAVGTQRT 240
QY	241	LALSHNNRYATGKVEYFRPVSEFTCDRSGSSTGHWSEYVDKNQVVELPIVDSVAPR 300		241	LALSHNNRYATGKVEYFRPVSEFTCDRSGSSTGHWSEYVDKNQVVELPIVDSVAPR 300
DB	241	LALSHNNRYATGKVEYFRPVSEFTCDRSGSSTGHWSEYVDKNQVVELPIVDSVAPR 300		241	LALSHNNRYATGKVEYFRPVSEFTCDRSGSSTGHWSEYVDKNQVVELPIVDSVAPR 300

Qy	301	PRYLPLAIPEDLADLVYVHGDPAVMWV	SOYKYILIRQPMLEKIEEATKKGKHNVI	360
Db	301	PRYLPLAIPEDLADLVYVHGDPAVMWV	SOYKYILIRQPMLEKIEEATKKGKHNVI	360
Qy	361	GVHVRRTKVGAEAFHFI	IEETVHVEEDFOLLARMOVDKKRYLADDDPALLKEATK	420
Db	361	GVHVRRTKVGAEAFHFI	IEETVHVEEDFOLLARMOVDKKRYLADDDPALLKEATK	420
Qy	421	YPSYEFISDNI	ISMSAGLIHNRTEISLGVILDIHFLSQADELVCTESSOVCRAVEITMO	480
Db	421	YPSYEFISDNI	ISMSAGLIHNRTEISLGVILDIHFLSQADELVCTESSOVCRAVEITMO	480
Qy	481	ALHPASANFRSLDDIYYEFG	PGPNAHNOIAIYPHORTEGEIIMPEBGDITIGVAGNHWDGYP	540
Db	481	ALHPASANFRSLDDIYYEFG	PGPNAHNOIAIYPHORTEGEIIMPEBGDITIGVAGNHWDGYP	540
Qy	541	KGVNRRKLGRTGLPSYKV	REKIEIETVKYPTYPEADK	575
Db	541	KGVNRRKLGRTGLPSYKV	REKIEIETVKYPTYPEADK	575

## RESULT 2

hypothetical protein C10F3.6 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence,revision 29-Oct-1999 #text,change 29-Oct-1999  
C:Accession: J121154  
C:Davidson, S.; Wohlmann, P.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid C10F3.  
A:Reference number: 221127  
A:Accession: J121154  
A:Status: preliminary: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-818 <DAV>  
A:Cross-references: EMBL:AF022968; PIDN:AA869886.1; GSFDB:GN00023; CESP:C10F3.6  
A:Experimental source: strain Bristol N2; clone C10F3  
C:Genetics:  
A:Gene: CESP:C10F3.6  
A:Map position: 5  
A:Introns: 30/1; 72/2; 235/2; 258/3; 285/2; 315/3; 457/3; 491/1; 544/3; 556/1; 626/1; 644/1

[illegible]

Oy	462	KRVATATDDPALLIEAKRKYSPYEISDNSTSMWAGJHNRTENSJCVLLIDHIFLSQAD	461
Dd	389	RRIPLASDPPYVPEKARDYPRNEYEGSTELAKTRAPOLNNRRITDSLMSGVITDIILSKVN	448
Oy	462	FLVCTFSSOVCRAVEIEMQLHPDASANFRSLDDIYFFGGFNAINQALAIYHQPTREGEI	521
Dd	449	YLCTCFSSOVCRMGLELPQSGADGSKFSHLDDIYFFGGQAHEVIATEDHIAONNKEI	508
Oy	522	PMEPEDDIGVAGNHWDGYPKGVNRKLGRGLTGLPYSPKYVR	559
Dd	509	DLKVGDKGVIAGNHNGYSKGTNRQTYKEGVFPSPKIR	546

### RESULT 3

hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86250  
R:Thellogiis: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizart, L.  
Native 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzita Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: H86250  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1025 <STO>  
A:Cross-references: GB:AE005172; NID:g4835785; PIDN:AAD30251.1; GSPDB:GN00141  
A:Genetics:  
A:Map position: 1

[illegible]



Db 736 NLEFGANAGKAMTFADKATTVSPYAKEVAGNSVIAHLKFKHGIINGIDPIDMDPYND 795  
OY 330 QFVYLLIRPOWLEKEI-----EEATKLGFK---HPYIGVIVRTDKVGAARHP 378  
Db 796 NFI-----PVYTSBENVVEGRKAKEELONRGKLSADFPVVGIIITRLTHQGHILKHA 850  
OY 379 I-----EEYTVHVEEDFQOLLARRMQV---DKRKYVLTATDPPALLKFAKTKY 421  
Db 851 IWRTERNGOYVLLGASADPRQNDPFLNLANGLHSHSGDRARLVLTDEPL-----SHLTY 906  
OY 422 PSYEFISDNISWSAGL 438  
Db 907 AGADFILVPSIFEPCGL 923

RESULT 4  
F89692  
cardamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: F89692  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F89692  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1057 <KUR>  
A:Cross-references: GB:BA000018; PID:q13701002; PIDN:BAB42298.1; GSPDB:GN00149  
A:Experimental source: strain N315  
A:Genetics:  
C:Gene: PYRAB  
C:Superfamily: cardamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 4.0%; Score 123.5; DB 2; Length 1057;  
Best Local Similarity 18.9%; Pred. No. 2;  
Matches 115; Conservative 85; Mismatches 221; Indels 189; Gaps 25;  
OY 43 LSKILAKLERLKOEDLRMMAGS-----LRIPGCPID-----OGP 78  
Db 318 IAKLAIAVGLTDEMNPITGSTYAFPTLDYVISKIRPFRDKEKREGLGTQMK 377  
OY 79 ASGRVAL-----EEOFMKAKEOIEI--YKQTKNGPKGDHETLRRIENGAKELMFLQIS 132  
Db 378 ATGEVMAIGRTYEEISLKAIRSLLEYGVNHLGLPNCESFDLDYIKERISHODDERLFFIGE 437  
OY 133 ELKRLKNEGNELOHNADEF-----LSDGHNERSIMTDLUYLSQTDGAGDMKEKAK 185  
Db 438 AIRRTTLEELHNMTQIDYFLHKFQNTIDIEHOLKEHOGDELYLKVAKDYG-FSDKITA 496  
OY 186 DLTELQORITVYLQNPCKDSKAKLVNINCKGCGCOLHNHVYCFMAYGTQRTLALES 245  
Db 497 HRFNMTTEEYVQLRBNNDIKPYKKV----- 522  
OY 246 HNMRYATGCMETVFRPVSESTDRSGSGTGHMSGEVKDKNVOV---ELPIYDSVHPRP 301  
Db 523 -----DTCAAEFESSTPYGYGTETENESIVYDKKELIIVLGSG--- 560  
OY 302 PYLPLAVPEDLADRLVRYHGDPAVWVSO--FVKYLLIRPO-----WLEKEI 346  
Db 561 ---PRTGCGVEFYATVH---AVMAIQAGYEAITVNNNPVTSTDSISDKLTFEPLT 614  
OY 347 EATKRLGFKHPVIGVNRRTDKVGAEAHPRIEYTVH-----VEEDF 390  
Db 615 EEDVANNIINLEKPKGVV---QFGGQTAINLADKLAKHGVKILGTSLLENLRADRKFE 670  
OY 351 QLLARRMVDKRYVLAID-DEALLKEAKTKY-----PSYEF-----ISDMSWSAGL 438  
Db 671 EALLRKINVPQGSATSPSEALANMAEIGYVVVYRPSYVYLGGRAMELVDN---DKEL 726

OY 439 HNRYTE-----NSLRGVILDIHFLSQADFLVCTFSQVCRAVEIMQALHPDASANFRSLD 494  
Db 727 ENYITQAKKASEHNVLD-RITCKELEVDNACIGGEVITITGIMEHIER----- 775  
OY 495 DIYYFGGNANQAIYHPQRTPEGE-----IPMEPG-DITGVAGNHW---DG-YP 540  
Db 776 ---AGVHSGSIANVPPQTLTEDELATLEDYTIKLAKGLNITIGLINIQFIADGVYV 830  
OY 541 KGVNRKLGRT 550  
Db 831 LEVNPRESSRT 840

RESULT 5  
D96796  
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96796  
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: D96796  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1871 <STO>  
A:Cross-references: GB:AE005173; NID:96143906; PIDN:AF04452.1; GSPDB:GN00141  
A:Genetics:  
C:Gene: F28016.15  
A:Map position: 1

Query Match 4.0%; Score 123; DB 2; Length 1871;  
Best Local Similarity 25.0%; Pred. No. 4.7;  
Matches 47; Conservative 40; Mismatches 73; Indels 28; Gaps 9;  
OY 24 YIGHLVNDHNSDSSNELSK-ILAKLERLKOEDLRMMAGSIRIPGCPIDGCPASGR 82  
Db 1065 HTGG-----EDNDHKKEEOKENVIAKAE-LNTEEDSFKKVEIEIKODHGLKRSVQAK 1118  
OY 83 VAALEE---OPMAKKEOIEIENYKQTKNG-----PGKHEI---LRRRIENGAKELMFF 129  
Db 1119 RQTEFEKQTRAMEKNYER-RKQTKDSLSKLAKGEPELGCHGRKCEEDRIEL--- 1174  
OY 130 LQSEL---KKIKLNEGNELOHNADEFSLDGNHRSITFDLYLSQTDGAGDMKEKAK 185  
Db 1175 VETETSDHKEKKKDEVDILRSQDTGKVDGERRRRSKORKIKHSVEDEIDQDDEDAE 1234  
OY 186 DLTELQOR 193  
Db 1235 EAAAVVSR 1242

RESULT 6  
S22695  
restin - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 05-Nov-1999  
C:Accession: S22695; S19853  
R:Biibe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asseibergs, F.A.M.; Cerletti, N  
EMBO J. 11, 2103-2113, 1992  
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in  
A:Reference number: S22695; MUID:92289675  
A:Accession: S22695  
A:Molecule type: mRNA  
A:Residues: 1-1427 <BIL>





A:Molecule type: DNA  
 A:Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>  
 A:Cross-references: EMBL:X08065; NID:96785; PIDN:CAA30854.1; PID:96786  
 R:Karrn, J.; Brenner, S.; Barnett, L.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983  
 A:Title: Protein structural domains in the *Caenorhabditis elegans* unc-54 myosin heavy chain  
 A:Reference number: A93958; MUID:83273600  
 A:Accession: A02993  
 C:Genetics:  
 A:Molecule type: DNA  
 A:Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473, 'G', 475-576  
 C:Genetics:  
 A:Gene: myo-1; CESP:R06C7.10  
 A:Map position: 1  
 A:introns: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co  
 F:177-184/Region: myosin motor domain homology <MOT>  
 F:177-184/Region: nucleotide-binding motif A (P-loop)  
 F:660-682/Region: actin binding #status predicted  
 F:764-778/Region: actin binding #status predicted  
 F:846-1938/Domain: coiled coil #status predicted <COI>  
 F:846-1160/Region: S2  
 F:1161-1938/Region: light meromyosin  
 F:128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
 F:183/Binding site: ATP (Lys) #status predicted  
 F:700,710/Active site: Cys #status predicted

Query Match 3.8%; Score 117; DB 1; Length 1938;  
 Best Local Similarity 20.5%; Pred. No. 13; Mismatches 139; Indels 168; Gaps 21;

Matches 97; Conservative 70; Mismatches 139; Indels 168; Gaps 21;

36 SDHSSRELKTIKLERLKQONEDLR---MAGSLRIPEGPIDGPASGR-----V 83  
 Db 1481 SRTSTVEFRKLRSMNLSQETLRREKKIFSQELRINEQITQCG--GRYQVEHKSV 1537  
 QY 84 RALEDFPMKAKEDIEYKQTKNG-PGKHDEILRRRIENGAKELMFLQSELKKNLBS 142  
 Db 1538 RLRLQEE---KDELQHALDEAEALAEESKVLK-----LQIEVOQIRSEIE 1580  
 QY 143 NELQRNADEFLSDGHNHR---SIMTDLVYLSTQDAGDMREKADITELVORRTIYQ 199  
 Db 1581 KRIOKEEDEFENRKRKHQRLSIOASL-----ELEAKSKELAK----- 1622  
 QY 200 NPKDCSKAKKLYCNKGGCGGCOLHNHVVYCFMIAVGTQRTALLESNNRY--ATGQWET 257  
 Db 1623 -----KKLETDIN-----QLEILDHANKANVDAQNLKK 1652  
 QY 258 VFRPVSETCTDRSGSTGHSCEVKKKNVQVELPIVDVYHRRPYL-----PLAYP 309  
 Db 1653 LRFQVREL-----QGQVDEQRREI---RENYLAERKRLALALS 1692  
 QY 310 EDLADRL-----VRHGDPAVMMVSGFVKYLIRPQLEKEIEE 348  
 Db 1693 EDLAHIEASDKKKQLEIEQALAKSSNTELGNNNA-----LSAKKRVENEVOQ 1743  
 QY 349 AFKKLGFKHPVIGVHVRPDYGAEA----AFHPLEETVHVE---EDFQLLRMYV- 399  
 Db 1744 ARNELDEYLNELKASERARKAADADRLAEVROEQEHVHVRQKSLNLAKELQAK 1803  
 QY 400 --DKRNVLTADDPALKEAKTKYPSYEFLSDNSISMSAGLHN--RYTENSLLG 449  
 Db 1804 IDDAERAMIOFGAKALAK-----VEDRVSLEALHSEORRHOESING 1846

RESULT 11  
 T32568  
 Hypothetical protein F37C4.6 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32568  
 R:Langston, Y.; Sansone, J.; Wohldmann, P.  
 Submitted to the EMBL Data Library, December 1997  
 A:Description: The sequence of C. elegans cosmid F37C4.

A:Reference number: Z21193  
 A:Accession: T32568  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-544 <LAN>  
 A:Cross-references: EMBL:AF036705; PIDN:AA95172.1; GSPDB:GN00022; CESP:F37C4.6  
 A:Experimental source: strain Bristol N2; clone F37C4  
 C:Genetics:  
 A:Gene: CESP:F37C4.6  
 A:Map position: 4  
 A:introns: 29/3; 230/2; 397/2; 484/3

Query Match 3.7%; Score 116; DB 2; Length 544;  
 Best Local Similarity 19.6%; Pred. No. 2.7;

Matches 110; Conservative 72; Mismatches 213; Indels 166; Gaps 28;

25 IGHILVRDNDHSHSRELKTIKLERLKQONEDLRRAAGSLRIPEGPIDGPASGRVR 84  
 Db 49 VGGAAVTEELVPGFRSRASYLSLIRPVVMDLNKKKGLRYHT-RNPNSFTP---IR 103  
 QY 85 ALDEDFPMKAKEDIEYKQTKNG-PGKHDEILRRRIENGAKELMFLQSELKKL 137  
 Db 104 NTHESLLGMDMAENKELAKSQRDAGNYRKEHFISE-----IVHSFEQL 150  
 QY 138 KNLGELQRNADEFLSDGHNHRISMTDLVYLSTQDAGDMREKAKDLTELVRRTIY 197  
 Db 151 MDYERPLDQKPI-----HKLLPHLYLLFKT--VQPLGLRNAVDVFEELMTAPISK 197  
 QY 198 LQN---PKDCSKK-----KLYCNKKGCGGCOLHNHVVYCFMIAVGTQRTALLES 246  
 Db 198 INMKWESVLUKATLTCTDGTGIGLAASPMDPGTGY-VLLNHYIG-----GLDEH 244  
 QY 247 N--WRATGQWETVFRPVSETCTDRSGSTGHSCEVKKKNVQVELP--IVDSVHRRP 302  
 Db 245 KQAMGYVYGGMGAVSNAIE-CAKSHCAEL-----YTQDQVQVELLDGNVAKGV- 293  
 QY 303 YLPLAVEDLADLVHVRHDDPAVMMVSGFVKYLIRPQLEKEIEBATKLGKHPVIGV 362  
 Db 294 --LSNGKELHSKIVSNATPHVTF-----NHLVKES-LPEFHNNINQIDVTSPTKI 344  
 QY 363 HVARTP-----KGAEAHPRIEETVHVE-EDFQL-----ARR----- 396  
 Db 345 NVAVKELPNEFLAKPNOGSEPM--PHNQTTIHNCENQVAVHDAVMYKNGRYSRRPIEM 402  
 QY 397 ---MQYDKRVYLATDPALKEAKTKYPSYEFLSDNSISMSAGLHNRYTENSLLG 453  
 Db 403 TTPSSVDRTIVDSADGHVLL-----FTQYTFPSKDGEMTEKKTEYAK----- 447  
 QY 454 IHFLSQADFLVCTESSQVCRAVEIMQALHPDASANF-----RSLDDIY----- 497  
 Db 448 -HVFSEIDAVAPNFSSSV--IGVDILTP--PDIQNFGITGCMIFHGSMSLDOLVYSRPI 502  
 QY 498 -----YFGGPMNH 505  
 Db 503 SKWSNTSTPLESLYLCGSGAH 523

RESULT 12  
 I48176  
 synaptonemal complex protein - golden hamster (fragment)  
 C:Species: *Mesocricetus auratus* (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
 C:Accession: I48176  
 R:Dobson, M.J.; Pearlmann, R.E.; Karlskaks, A.; Spyropoulos, B.; Moens, P.B.  
 J. Cell Sci. 107, 2749-2760, 1994  
 A:Title: Synaptonemal complex proteins: occurrence, epitope mapping and chromosome di  
 A:Reference number: I48176; MUID:95181577  
 A:Accession: I48176  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-845 <RES>  
 A:Cross-references: GB:I32978; NID:9598124; PIDN:AA42039.1; PID:9598125  
 C:Genetics:

A:Gene: syn1  
Query Match 3.7%: Score 114.5; DB 2; Length 845;  
Best Local Similarity 17.6%: Pred. No. 6.2; Mismatches 181; Indels 99; Gaps 16;  
Matches 78; Conservative 84; Mismatches 181; Indels 99; Gaps 16;  
OY 29 LVRDSDHSDHSSRLKILAKLERLKOONEDLRMRAGSLRIPEGPID--QGPASGRVAL 86  
D 275 MTKRRNNNEVLEELKTLAEDQKLDKQOYEKLAELQSGEDTLTLLQTRKEVINDL 334  
OY 87 EEOFKAKEQIENYKKQTKNGPKDHELRRIENGAKELMFELOSLEKLNK----- 139  
D 335 EEOFLVTKISQONSQV-----BELKTKLEE-----EKLNAELTASC 373  
OY 140 -----LEGLELORHDEFLSDLGHNERSIMT-----DLVYLSQTDGMDRREKAKLTLEY 191  
D 374 GKLSENKLLQOETNDMALEKTLQOEDITNSKQOERMLOQIENLE--KETHLRDELISV 432  
OY 192 QRRITVYQNPADK-----SKAKLVNINR-----GCG-----YGCOL 224  
D 433 RKEFIQOGENEYKCKLDSEENARSTIECEVLKKEKQMKTLKCNLKRQAEKSKYIEL 492  
OY 225 HHVYCFMAYGTOFTLALSHNMRYATGMEYFVRVS---ETCTDRSGSGTHWAGEV 281  
D 493 HQ-----ENKALKKSSAESKQOLNAYEIKVKKLQLESAKQKFOEKTNDYQKEI 542  
OY 282 KDKNQVVELPIVDSVHRRPYLPLAVPEDLADRLVVRHGDPAVWVWQFVYLLRPQW 341  
D 543 EVKTKIS--EEKKLGEVERAKAMVDEAVKLQKEIDLRQHK-----IAMEVALMEKHQ 594  
OY 342 LEKEIEATKRLKGFKNPIYGVHVRTRDKVGAFAHPIEETVHVEEFOQLARMQV-- 399  
D 595 YDKIVEERDSELGL-----CKNRQEQDLSVKT--LETELSINRLNLSLAKQLEIER 645  
OY 400 -DKKRVYLATDDPALKEAKTK 420  
D 646 EEKEKLKEKENTAILKDKDK 667  
RESULT 13  
889971  
conserved hypothetical protein SA1661 [Imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: B89971  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: B89971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-978 <KUR>  
A:Cross-references: GB:BA000018; PID:g13701636; PIDN:BA042929.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA1661  
Query Match 3.6%: Score 112; DB 2; Length 978;  
Best Local Similarity 20.3%: Pred. No. 11;  
Matches 107; Conservative 84; Mismatches 200; Indels 136; Gaps 27;  
OY 14 ILPAWGTLFYIGHLVBDN--DHSHSSRELKTLAKLERLKOONEDLRMRAGSLRIPE 71  
D 490 IIFAVLTIVFYIGIMSKSEVDYSEAITDEIEIKQALAIL--DENYDL-----D 538  
OY 72 GPIDGSPASGRVALBEQFMKAKBOIENYKKQTKNGPKD-----HELRRIENGAKELM 127  
D 539 FDLDE--QYRDRHMQDALKKDKLLE--KROYIEGRDLAKGRDELQSTVENKDELY 594  
OY 128 F-----FLOSLEKLNLEGNELORHDEFLSDLGHNERSIMTDL----- 167

D 595 LSKISNDLIVDSISTMANIKR-----LDQHSIDLNQOPOQLVQEDLPFYVNHAEVTKS 648  
OY 168 ---YI--LSQTDGADMRREKAKDL-----TELVRRIITVL-----ONPKDSCKA 207  
D 649 QFVYKTLSTLFHDYQWML--KSAEDTNEKRWIENAEYTKITLNLNOLNEENKEIT-- 705  
OY 208 KKLVCNINIKGGGCGQLNH---VYVCFMAYGTOFTLALSHNMRYATGMEYFVRVS 264  
D 706 -ALFDFIVNGTEEDFYQHNEEDYQTSNLSRPNDLTKYLENONYSLEUS-----SUSE 758  
OY 265 TCTDRSGSGTHWAGEVCKDNQVVELPIVDSVHRRPYLPLAVPEDLADRLVVRHGDPA 324  
D 759 KTTAQLSEEDHILATQVDEYNQYLEMQ-----AQVSDLSAQINNHETDTT 804  
OY 325 VWWVQFVYLLRPOPMLEKEIEEATK--KLGFNRYGVYVRR--TDKVGAEAFHPIE 380  
D 805 -----LANLRHEVHSLKQNLINDIAKDWASLSYLSQSLVDEHIKQDKRLPOVINEAVE 857  
OY 381 -----EYTV-----HVEDPQILARRMQYDKKRVYLATDDPALKEAKTKY 421  
D 858 ILKHLTDGRITMINEDSITVYKHVNGQLYDPELSOSTKELLVYAL--RISLTKLRPY 916  
OY 422 PSYEFISDSISWSAALHNRYTE--NSLRGYLIDHFLSQADFLVC 465  
D 917 PEPPLVDDAFVHFD---KKRTEKMLNLTLS--LSEHY--QVLYFTC 955

RESULT 14  
A70387  
conserved hypothetical protein aq\_1006 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 02-Feb-2001  
C:Accession: A70387  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: A70387  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-978 <AOF>  
A:Cross-references: GB:AE000718; NID:g2983504; PIDN:AC07092.1; PID:g2983515; GB:AE00  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: aq\_1006  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032  
Query Match 3.6%: Score 111.5; DB 2; Length 978;  
Best Local Similarity 19.4%: Pred. No. 12;  
Matches 78; Conservative 74; Mismatches 130; Indels 121; Gaps 18;  
OY 41 RELSKTLAKLERLKOONEDLRMRAGSLRIPEGPIDGSPASGRVALBEQFMKAKBOIENY 100  
D 336 KELEHRLKRLQEKIKELSELQSSSSLKKEKEEYDQ-----KQEFEDLSERVEKG 386  
OY 101 KK---QTKNGPKDHELRRIENGAK--ELMFLQSELKLNLEGNELORHDEFLS 154  
D 387 KKLVAETEKELEKIKELSEEEYTSLKMKERLVLQRLKELKEKEG--QLENLTQKYKE 445  
OY 155 DLGHNERSIMDLYLSQTDGMDRREKAKDLTELYQRRITVYQNPDCSAKKLVYCN 214  
D 446 KKKVHEK--VLNLEKLE-----RELKERLHANHVAAYL--SPGD-----TCV 488  
OY 215 NKGCYGGQOLNHVVCFMAYGTOFTLALSHNMRYATGMEYFVRVSSETCTDRSGSST 274  
D 489 ---CG-----GIYRGKALEN-----VDAEGISE 508  
OY 275 GHWGSEVCKDNQVVELPI-----VDSVHRRPYLPLAVPEDLADRL----- 316  
D 509 LKHAKELKEKEEREIDTTLTKLYAQKINSILKEEMERLNEVEBELRKEIPENILKEIRIKLEE 568



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 21:34:02 ; Search time 2009 Seconds  
(without alignments)  
13930.215 Million cell updates/sec

Title: US-09-839-136-1  
Perfect score: 1728  
Sequence: 1 atgcgcatactgactgcttc.....atcccgagcgtcagaatga 1728

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlun:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	46.0	1066	14	BM928521 AGENCOURT
2	772.4	44.7	937	14	BO892101 AGENCOURT
3	753.6	43.6	940	14	BO653329 AGENCOURT
4	741.6	42.9	886	14	BO423860 AGENCOURT
5	718.4	41.6	752	14	BO604588 MI-P-CPI-
6	691.6	40.0	834	9	AU124128 AU124128

7	619.6	35.9	965	13	B1546364
8	573	33.2	936	12	BE893107
9	567.6	32.8	769	12	BG501191
10	557.8	32.3	702	13	BG973321
11	557.2	32.2	715	12	BG771085
12	555.8	32.2	656	14	BO552954
13	552.4	32.0	900	14	BO731540
14	536.8	31.1	653	12	BG619839
15	513.2	29.7	623	14	BO305885
16	512.2	29.6	1273	13	BO720741
17	508	29.4	1002	13	B1517671
18	502.6	29.1	621	14	BO305594
19	496.4	28.7	579	13	B1964023
20	495.6	28.7	600	14	BO305888
21	492.8	28.5	635	14	BO307749
22	489.4	28.3	1009	14	BO223892
23	487.2	28.2	573	9	A1122135
24	484.2	28.0	584	14	BO305887
25	483.2	28.0	888	12	BE731842
26	482.2	27.9	607	14	BO307741
27	475	27.5	594	14	BO308365
28	469.6	27.2	732	10	AW635821
29	463.6	26.8	970	14	BO926824
30	462.4	26.8	555	14	BO309171
31	460.2	26.6	529	10	BE543978
32	455.2	26.3	563	14	BO307126
33	455	26.3	598	14	BO305867
34	451.4	26.1	538	14	BO305873
35	449.8	26.0	534	14	BO307345
36	448.2	25.9	534	10	BE690071
37	445.4	25.8	697	9	AJ398890
38	436.2	25.2	505	14	BM717378
39	420	24.3	511	14	BO306837
40	419.2	24.3	826	12	BF246161
41	419	24.2	631	14	BO309231
42	414.8	24.0	484	10	AW387766
43	414.4	24.0	499	14	BO305595
44	412.4	23.9	587	14	BO305859
45	412.2	23.9	458	9	AA057866

## ALIGNMENTS

RESULT 1  
BM928521  
LOCUS AGENCOURT\_6715254 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:5798262  
DEFINITION 5', mRNA sequence.  
ACCESSION BM928521  
VERSION BM928521.1 (GI:19378900)  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1066)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Place: LNCM2025 row: p column: 07  
High quality sequence stop: 666.  
Location/Qualifiers

## FEATURES

source 1..1066

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5798262"
/clone.lib="NIH_MGC.100"
/issue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT      315 a      235 c      248 g      264 t      4 others
ORIGIN
Query Match      46.0%; Score 795; DB 14; Length 1066;
Best Local Similarity 89.8%; Pred. No. 5.6e-188;
Matches 885; Conservative 0; Mismatches 98; Indels 3; Gaps 3;
QY 366 AGCTAAGACCTGCTGTTTCTTACAAAGTTCAGAAATTAAGAAATTTAGAGG 425
    |||||||
Db 1 AGCTAAGACCTGCTGTTTCTTACAGAGTGAATTAAGAAATTTAGAGG 60
QY 426 AATGAACTCCAAGACATGATGATTAATTCAGATTGGACATCATGAAGTGC 485
    |||||||
Db 61 AATGAACTCCAAGACATGATGATTAATTCAGATTGGACATCATGAAGTGC 120
QY 486 TATTAATGACGATCTATCTACTAGTCAAGATGGGAGGTGATGGCGTGA 545
    |||||||
Db 121 TATTAATGACGATCTATCTACTAGTCAAGATGGGAGGTGATGGCGGAAA 180
QY 546 GAGGGCCAAAGATCTAGACAGTGTGCCAGCGGAATAACATATCTTACAGATCCCA 605
    |||||||
Db 181 AGAGGCCAAAGATCTAGACAGTGTGCCAGCGGAATAACATATCTTACAGATCCCA 240
QY 606 GGAATGACGAAAGCCAAAGAGTGTATATCAACAAGGCTGTGCTATGGCTG 665
    |||||||
Db 241 GGAATGACGAAAGCCAAAGAGTGTATATCAACAAGGCTGTGCTATGGCTG 300
QY 666 TCAGCTTCATCATGTAGTGTACTGTTTATGATTCATATGCAACCCAGGAACTGCG 725
    |||||||
Db 301 TCAGCTTCATCATGTAGTGTACTGTTTATGATTCATATGCAACCCAGGAACTGCG 360
QY 726 CTTGGATCTCAAAATTTGGGCTAGCTGTGGGGAATGGGAAAGTGTGTTAGACCTGT 785
    |||||||
Db 361 CTTGGATCTCAAAATTTGGGCTAGCTGTGGGGAATGGGAAAGTGTATTTAGGCTGT 420
QY 786 AAGTGAGACGTGCAGACAGATCTGGCAGCTCCAGTGCATTTGGTCAAGTAA 845
    |||||||
Db 421 AAGTGAGACGTGCAGACAGATCTGGCAGCTCCAGTGCATTTGGTCAAGTAA 480
QY 846 GGACAAAATGTTCAAGTGTGAGCTCCCATTTGAGACAGTGTTCCTGCTCTCC 905
    |||||||
Db 481 GGACAAAATGTTCAAGTGTGAGCTCCCATTTGAGACAGTGTTCCTGCTCTCC 540
QY 906 ATATTATACCCCTGGCTGTCCCAAGACCTTGCAGATCGACTGTACAGATCCATGCTGA 965
    |||||||
Db 541 ATATTATACCCCTGGCTGTACCAAGACCTTGCAGATCGACTGTATTTAGTGTGA 600
QY 966 TCTGCGAGTGTGGGTATCCAGATTTCAGATTAATTTGGCGCCACAAACCTGGCT 1025
    |||||||
Db 601 CCTGCGAGTGTGGGTATCCAGATTTCAGATTAATTTGGCGCCACAAACCTGGCT 660
QY 1026 GGAAGAAAGAAATAGAAAGAGCCACAAAGAGCTAGGCTTCAAAATTCAGATTATTTGGAGT 1085
    |||||||
Db 661 AGAAANAGAAATAGAAAGAGCCACAAAGAGCTTGGCTTCAAAATTCAGATTATTTGGAGT 720
QY 1086 CCATGTGTAAGCGACAGCAAAAGTGGGAGCGGAGACGCTTCCATTCCTTTAGAGATA 1145
    |||||||
Db 721 CCATGTGTAAGCGACAGCAAAAGTGGGAGCGGAGACGCTTCCATTCCTTTAGAGATA 780
```

```
QY 1146 CACGGTGCAGCTTGAAGAAGACTTTCAGCTTCTTGGTCGACAGATGAGTGAATAAAA 1205
    |||||||
Db 781 CATGTGTCATGTTGAGAGAAATTTTCAGCTTCTGTCAGCGAGATGCMGTGACACANMA 840
QY 1206 AAGGCTGATATTTGCCACAGATGACCTGCTTTGTTAAAGAGCAAAAACAAGTACCC 1265
    |||||||
Db 841 AAGAGTATATTTGGCCACAGATGACCTTCTTATTTAAAGAGCAAAAACAAGTACCC 900
QY 1266 CAGTTTGAATTT-ATTAGATTAAGTCTATCT-CTTGGTCAGCTGGACTACATATTCG- 1322
    |||||||
Db 901 CAAATTGATTTAATTAATGATTAATCTTATTTCCCTGGTCAGCGGACTTGCCTATCGN 960
QY 1323 ATATACGAAATTTCACTTCGGGGT 1348
    |||||||
Db 961 ATACCCGGAATAATTTCCCTTCGGGG 986

RESULT 2
B0892101
LOCUS
DEFINITION
B0892101 937 bp mRNA linear EST 16-AUG-2002
AGENCOURT 8743832 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6389725
5', mRNA sequence.
ACCESSION
B0892101
VERSION
B0892101.1 GI:22284115
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus.
REFERENCE
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 937)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M13875 row: d column: 14
High quality sequence start: 130
High quality sequence stop: 780.
Location/Qualifiers
1. 937
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6389725"
/clone.lib="NIH_MGC.129"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: This
is a NIH_MGC Library."
BASE COUNT      268 a      215 c      209 g      244 t      1 others
ORIGIN
Query Match      44.7%; Score 772.4; DB 14; Length 937;
Best Local Similarity 90.0%; Pred. No. 2.4e-182;
Matches 838; Conservative 0; Mismatches 92; Indels 1; Gaps 1;
QY 736 CACAATTTGGCGCTAGCTACTGTGGGATGGGAAGCTGTGTTGACCTGTAAGTGAGAGC 795
    |||||||
Db 1 CACAATTTGGCGCTATGCTGCTGATGGAGACTGTGTTGACCTGTAAGTGAGAGCA 60
QY 796 TGCACAGACAGATCTGCGACCTCCACTGACATTTGGTCAGCTGAAGTAAAGCAAAAAT 855
    |||||||
Db 61 TGTACAGACAGATCTGCGCTCCACTGACACTGTGTCAGTGAAGTAAATGACAAAAC 120
```





Db 541 ATATTACCTTGGCTGTACACAGACCTCCAGATCGACTTGTACGATGCGATGCTGA 600  
Qy 966 TCCGTGAGTGTGGTATCCAGTTCATGACTGATTCGCCACACCTGGCT 1025  
Db 601 CCGTGAAGTGTGGTGTCTGATGTCATATGATTCGCGCCACAGCTTGCT 660  
Qy 1026 GGAAGAGAAATAGAGAGGCCACCAAGAGCTTAAACATTCAGTATTTGAGT 1085  
Db 661 AGAAAAAGAAATAGAGAGGCCACCAAGAGCTTAAACATTCAGTATTTGAGT 719  
Qy 1086 CCAATGTAGACGACAGACAAAGTGGAGGAGGAGCCCTTCATCCATTCAGAGATA 1145  
Db 720 CCAATGTAGACGACAGACAAAGTGGAGAGAGAGTCCCTTCATTCATTCAGAGATA 779  
Qy 1146 CAGCGT-GCAGCTTGAAGAACTTCAGCTTCTGCTCCAGATGCAAGTGGATAAA 1204  
Db 780 CAGTGTGAGTGTGAAGAACTTCAGCTTCTGCTCCAGATGCAAGTGGATAAA 839  
Qy 1205 AAAGGCTGA-TTTGGCCACAGATGACCTTCTTGT-7AAAGAGCAAAACAAAGTA 1262  
Db 840 AAAGAGTGTATTTGGCCACAGATGACCTTCTTGTATTAAGAGCAAAACAAAGTA 899  
Qy 1263 CCCCAGTTATGAA-TTTATAGTATTAATCTATCTTGT 1301  
Db 900 CCCCATTATGATTTTATATGATTAATCTATCTTGT 939

RESULT 4  
BO423860 886 bp mRNA linear EST 23-MAY-2002  
LOCUS BO423860  
DEFINITION AGENCOURT\_7894821 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6158200  
5', mRNA sequence.

ACCESSION BO423860  
VERSION BO423860.1 GI:21119175  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 886)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@pds-remail.nih.gov  
Tissue Procurement: ATCC/DCPD/BTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM13505 row: e column: 17  
High quality sequence stop: 641.

FEATURES  
source  
1.886  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6158200"  
/clone\_lib="NIH-MGC-72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 273 a 185 c 212 g 215 t 1 others  
ORIGIN

Query Match 42.9%; Score 741.6; DB 14; Length 886;  
Best Local Similarity 91.2%; Pred. No. 1.2e-174;  
Matches 798; Conservative 0; Mismatches 75; Indels 2; Gaps 1;

Qy 408 ATTAAGAAATTTAGAGAAATGAACTCCAAAGACATGCAATGTAATTTCTACATTT 467  
Db 1 ATTAAGAAATTTAGAGAAATGAAATGAACTCCAAAGACATGCAATGTAATTTCTTGGATTT 60  
Qy 468 GGGACATCATGAAAGGCTATATGAGGAGTCTATCTACCTACCTAGTCAAAAGATGGGG 527  
Db 61 AGGACATCATGAAAGGCTATATGAGGAGTCTATCTACCTAGTCAAAAGATGGGG 120  
Qy 528 AGGTGATTTGGCGTGAAGAGAGGCCAAAGATCTGACAGAGTGGTCCAGCGAGAAATAC 587  
Db 121 AGGTGATTTGGCGGGAAGAAAGAGGCCAAAGATCTGACAGAACTGGTTCCAGCGAGAAATAC 180  
Qy 588 ATATCTTCAGAAATCCCAAGAGCTGACAGCAAGCCAAAGAAAGCTGGTGTGAATATCAACA 240  
Db 181 ATATCTTCAGAAATCCCAAGAGCTGACAGCAAGCCAAAGAAAGCTGGTGTGAATATCAACA 240  
Qy 648 AGGCTGGGTATGGGCTGACAGCTCCATCATGATGATGATGATGATGATGATGATGATG 707  
Db 241 AGGCTGGGTATGGGCTGACAGCTCCATCATGATGATGATGATGATGATGATGATGATG 300  
Qy 708 CACCCAGCAACACTGCGCTTGGAAATCTCAATTTGGCGTACCTTACCTGAGGAGTGGGA 767  
Db 301 CACCCAGCAACACTGCTATCTTGGAAATCTCAATTTGGCGTACCTTACCTGAGGAGTGGGA 360  
Qy 768 AACTGTGTTAGACCTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 827  
Db 361 GACTGTATTTAGGCTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420  
Qy 828 TTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 887  
Db 421 CTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 480  
Qy 888 TGTTCATCTGCTGCTGCATATTTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 947  
Db 481 TGTTCATCTGCTGCTGCATATTTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
Qy 948 TGTTCATCTGCTGCTGCATATTTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007  
Db 541 TGTTCATCTGCTGCTGCATATTTTACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
Qy 1008 TCCGCCACAACTGCTGCTGGAAGAAATAGAAAGAGGCCACCAAGAGCTTACCTTCAA 1067  
Db 601 CCGCCACAGACCTTGGCTGAGAAAAAGAAATAGAAAGAGGCCACCAAGAGCTTACCTTCAA 660  
Qy 1068 ACATTCAGTATTTGAGTGCATATTTAGACGACAGCAAAAGTGGAGCGAGACCTT 1127  
Db 661 ACATTCAGTATTTGAGTGCATATTTAGACGACGACAGCAAAAGTGGAGCGAGACCTT 720  
Qy 1128 CCAATCCATTTGAGTATACAGCGGTGACAGCTTGAAGAAAGTTCACCTTCTGCTGCGAG 1187  
Db 721 CCAATCCATTTGAGTATACAGCGGTGACAGCTTGAAGAAACATTTTACCTTCTGCGAG 780  
Qy 1188 AATGCAAGTGTATTAAGAAAGAGGTGATTTGGCCACAGATGAGCC--TTCCTTTGTTAAA 1245  
Db 781 AATGCAAGTGTATTAAGAAAGAGGTGATTTGGCCACAGATGAGCCCTTCTTTATTAAG 840

RESULT 5  
BO604588 752 bp mRNA linear EST 24-JUN-2002  
LOCUS BO604588  
DEFINITION MI-P-CPI-nzc-f-18-0-UI-s1 MI-P-CPI Sus scrofa cDNA clone  
MI-P-CPI-nzc-f-18-0-UI 3', mRNA sequence.

ACCESSION BO604588  
VERSION BO604588.1 GI:21551314  
KEYWORDS EST.  
SOURCE Sus scrofa pig.  
ORGANISM Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.



Best Local Similarity 90.7%; Pred. No. 3.8e-162;  
Matches 747; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 905 CATATTACCCTGGCTGCCAGAAAGACCTTGCGATGCGATTGACAGTCCATGGTG 964  
DB 1 CATATTACCCTGGCTGCCAGAAAGACCTTGCGATGCGATTGACAGTCCATGGTG 60

QY 965 ATCTGCTAGTGTGGGTATCCAGTTTGTCAAGTACTGTGATGCGCCCAACCTGGC 1024  
DB 1 ATCTGCTAGTGTGGGTATCCAGTTTGTCAAGTACTGTGATGCGCCCAACCTGGC 1024

QY 61 ACCCTGCACTGTGGNGGTGTCTCAGTTTGTCAATACTGTGATCGGCCACAGCCTGGC 120  
DB 1 ACCCTGCACTGTGGNGGTGTCTCAGTTTGTCAATACTGTGATCGGCCACAGCCTGGC 120

QY 1025 TGGAAAGGAATAGAGAGGCCACCAGACCTAGCGCTTCAACATCCAGTTATGGAG 1084  
DB 121 TGGAAAGGAATAGAGAGGCCACCAGACCTAGCGCTTCAACATCCAGTTATGGAG 180

QY 1085 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGCCAGCTCCATCCCATTTGAGGAT 1144  
DB 181 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGCCAGCTCCATCCCATTTGAGGAT 240

QY 1145 ACACGGTGCAGCTTGAAGAAGACTTTCACCTTCTGCTCGAGAAATGCAAGTGATAAA 1204  
DB 241 ACATGTTGATGTGTGAAGAACATTTTACCTTCTGCTCGAGAAATGCAAGTGATAAA 300

QY 1205 AAGGCTGATTTTGGCCACAGATGACCTTCTTTTAAAGAGGCAAAACAAAGTACC 1264  
DB 301 AAGGCTGATTTTGGCCACAGATGACCTTCTTTTAAAGAGGCAAAACAAAGTACC 360

QY 1265 CCAGTATATATTTTATGATGATCTATCTCTGTCAGCTGATGATGATGATGAT 1324  
DB 361 CCATTTATATATTTTATGATGATCTATCTCTGTCAGCTGATGATGATGATGAT 420

QY 1325 ATACAGAAATTCACCTTCCGGGTGTGATCTGATATACATCTTCTCCAGGACACT 1384  
DB 421 ATACAGAAATTCACCTTCCGGGTGTGATCTGATATACATCTTCTCCAGGACACT 480

QY 1385 TCTTATGTTGATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1444  
DB 481 TCTTATGTTGATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 540

QY 1445 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGCCAGCTCCATCCCATTTGAGGAT 1504  
DB 541 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGCCAGCTCCATCCCATTTGAGGAT 600

QY 1505 CAATGCGCCAGCAACCAATTTGCATTTATCTCACCAACCTGGAAGTGAAGAAATCC 1564  
DB 601 AGAATGCCACATCAATTTGCATTTATGTCACCAACCTGGAAGTGAAGAAATTC 660

QY 1565 CCATGGAACCTGGAGATATTTATTTGGTGGTGAATACATGAGTATCTTAAG 1624  
DB 661 CCATGGAACCTGGAGATATTTATTTGGTGGTGAATACATGAGTATCTTAAG 720

QY 1625 GTGTTAAAGAAAGTGGGAGGAGGCGCTATATCCCTCTACAAAGTTGAGAGAA 1684  
DB 721 GTGTTAAAGAAAGTGGGAGGAGGCGCTATATCTCTCTACAAAGTTGAGAGAA 779

QY 1685 TAGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAGTAA 1728  
DB 780 TAGAAACAGTCAAGTACCCACATATCTGAGGCTGACAAGTAA 823

RESULT 7  
LOCUS B1546364 965 bp mRNA linear EST 05-SEP-2001  
DEFINITION 60318856f1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5260284 5',  
mRNA sequence.  
ACCESSION B1546364  
VERSION B1546364.1 GI:15433676  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 965)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palokvits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihaki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
Plate: LLM11655 row: p column: 13  
High quality sequence stop: 828.

FEATURES  
source  
location/Qualifiers  
1..965  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5260284"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.5 kb and  
normalized for R0F 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI/NHRI, National  
Institutes of Health). Note: this is a NIH-MGC Library."  
BASE COUNT 298 a 213 c 207 g 247 t  
ORIGIN

Query Match 35.9%; Score 619.6; DB 13; Length 965;  
Best Local Similarity 91.2%; Pred. No. 4e-144;  
Matches 680; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

QY 973 GTGTGGTGGGTGCCAGTTGTCAAGTATGATTCGCCCAACCTGGTGAAG 1032  
DB 4 GTGTGGTGGGTGTCTCAGTTTGTCAAAATGATTCGCCCAACCTGGTGAAGAA 63

QY 1033 GAAATAGAAAGGCCACCAAGAGCTAGGCTTCAACATCCAGTTATGAGTCCATGTT 1092  
DB 64 GAAATAGAAAGGCCACCAAGAGCTTGGCTTCAACATCCAGTTATGAGTCCATGTC 123

QY 1093 AGAGCCACAGACAAAGTGGAGCGAGACGCTTCCATTCCTTGGAGATACACGCTG 1152  
DB 124 AGAGCCACAGACAAAGTGGAGCGAGACGCTTCCATTCCTTGGAGATACATGCTG 183

QY 1153 CAGTGTGAAGAGCTTTCAGCTTCTGTGCGACAAATGCAAGTGAATAAAAAGGGTG 1212  
DB 184 CAGTGTGAAGAGCTTTCAGCTTCTGTGCGACAAATGCAAGTGAATAAAAAGAGTG 243

QY 1213 TATTTGGCCAGATGACCTGCTTTGTTAAAGAGGCAAAACAAATGACCCAGTTAT 1272  
DB 244 TATTTGGCCAGATGACCTGCTTTGTTAAAGAGGCAAAACAAATGACCCAGTTAT 303

QY 1273 CAATTTATTTAGTGAATCTATCTTCTGTGAGCTGAGCTACATATCGATATACAGAA 1332  
DB 304 CAATTTATTTAGTGAATCTATCTTCTGTGAGCTGAGCTACATATCGATATACAGAA 363

QY 1333 AATTCAGTTCGGGGTGTGATCTGATTTACACTTTCCTCCAGCAGCAGACTTCTAGTG 1392  
DB 364 AATTCAGTTCGGGGTGTGATCTGATTTACACTTTCCTCCAGCAGCAGACTTCTAGTG 423

QY 1393 TGTACTTTTTCATCCAGGTGTGATGATGCTTTATGAATCATGCAAGCCTGATCT 1452  
DB 424 TGTACTTTTTCATCCAGGTGTGATGATGCTTTATGAATCATGCAAGCCTGATCT 483

QY 1453 GATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTATTTTGGAGGCCCAATGCC 1512  
DB 1453 GATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTATTTTGGAGGCCCAATGCC 1512

Db 484 GATGCTCTGCAAACTTCATCTTTAGATGACATCTACTATTTTGGGGCCAGAAATGCC 543

Qy 1513 CACAACCAATTTGCAATTTNCTCCACCAACCTGCAACTGAGAGAAATCCCATGGAA 1572

Db 544 CACATCAAAATTTGCAATTTNCTCCACCAACCTGCAACTGAGAGAAATTTCCATGGAA 603

Qy 1573 CCTGGAGATATTTATTTGGTGTGGCTGGAATTCACAGTGGCTATTCCTAAAGTGTAAAC 1632

Db 604 CCTGGAGATATTTATTTGGTGTGGCTGGAATTCATTTGGATGGCTATTTCTAAAGGTGTCAAC 663

Qy 1633 AGAAATCGGAGAGAGAGGCGCTATATCCCTCC-TACAAGCTTGAGAGAGATTA-GAAA 1650

Db 664 AGGAAATTTGGAGAGAGGCGCTATATCCCTCTTACAAAGTTCGAGAGAGATGACGAA 723

Qy 1691 CAGTCAGTACCCCAATATCCCGAG 1716

Db 724 CGGTCAAGTACCCCAATATCCGTGAG 749

RESULT 8  
BE893107 936 bp mRNA linear EST 20-OCT-2000  
LOCUS 601437002F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3922104 5',  
DEFINITION mRNA sequence.  
ACCESSION BE893107  
VERSION BE893107.1 GI:10354133  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 936)  
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
plate: LLM8756 row: C column: 01  
High quality sequence start: 3  
High quality sequence stop: 696.  
Location/Qualifiers  
1. .936  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3922104"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site\_1: Not;  
Site\_2: Salt; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 260 a 202 c 244 g 230 t

ORIGIN

Query Match 33.2%; Score 573; DB 12; Length 936;  
Best Local Similarity 89.88; Pred. No. 1.8e-132;  
Matches 660; Conservative 0; Mismatches 70; Indels 5; Gaps 4;

Qy 351 GAGGATTTGAAATGAGACTCTGCTTTTCTACAAAGTGAAGTGAAGAAATT 410

Db 8 GTGGCTTGAAGATGAGACTCTGCTTTTCTACAGAGTGAATTTGAAGAAATT 67

Qy 411 AAGAAATTTGAAGAAATCACTCAAGAGCATGAGATGAATTTCTATCAGATTGGG 470

Db 68 AAGAACTTAGAAGAAATCACTCAAGAGCATGAGATGAATTTCTTTTGGATTAGG 127

Qy 471 ACATCATGAAGGCTATTAATGAGGATCTTACTACTGATGCAACAGATGGGCGAG 530

Db 128 ACATCATGAAGGCTATTAATGAGGATCTTACTACTGATGCAACAGATGGGCGAG 187

Qy 531 TGATTTGGCGTGAAGAGAGGCGCAAGATCTGACAGAGCTGGTCCAGGAGAAATACATA 590

Db 188 TGATTTGGCGGGAAGAGAGGCGCAAGATCTGACAGAGCTGGTCCAGGAGAAATACATA 247

Qy 591 TCTTCAGATTCACAGAGACTGACAGCAAGGCCAAGAGCTACTGTGTATATCAACAAAG 650

Db 248 TCTTCAGATTCACAGAGACTGACAGCAAGGCCAAGAGCTACTGTGTATATCAACAAAG 307

Qy 651 CTGTGGCTATGGCTGTACAGCTCCATCATGTAGTGTACTGCTTATGATGTGATGGCAC 710

Db 308 CTGTGGCTATGGCTGTACAGCTCCATCATGTAGTGTACTGCTTATGATGTGATGGCAC 367

Qy 711 CCAGCAACACTCCGCTGGAAATCTCAAAATTTGGCGTACGCTACTGGGAGATGGAAAC 770

Db 368 CCAGCAACACTCATCTTTGGAATCTCAGAAATTTGGCGTATCTACTGTGATGGAGAC 427

Qy 771 TGTGTTAGACCTGTAAAGTGAAGAGAGAGAGAGAGATCTGGCAGCTCCACTGACATTG 830

Db 428 TGTATTTAGGCTGTAAAGTGAAGAGAGAGAGAGAGATCTGGCAGCTCCACTGACACTG 487

Qy 831 GTCAAGTGAAGTGAAGAGCAAAATGTTACAGCTGGTGGAGTCCCATGTTAGACAGCT 890

Db 488 GTCAAGTGAAGTGAAGAGCAAAATGTTCAAGTGGTCCCATGTTAGACAGCT 547

Qy 891 TCATCTCTGCTCCATATTTACCCCTGGCTGCTCCAG-AGAGCTTGACAGATGACTTG 949

Db 548 TCATCTCCCGT-CTCATATTTTACCCCTGGCTGCTGTCACGAAAGAGCTGGCAGATGACTTG 606

Qy 950 TACGATTCATG--TGATCTGACAGTGTGGTGTATTCGAGTTGTCAACTTATGAT 1007

Db 607 TACGATTCATGAGGCTGACCTGACAGTGTGGTGTCTGAGTTGTCAAACTTATGGA 666

Qy 1008 TCGGCCCAACCCCGCTGGGAAAGAAATGAAAGAGGCGCAAGAGCTTACCTTCAA 1067

Db 667 TCGGCCAC-ACCTTGTTAGAAAGAAATGAAAGAGGAGTCAAGATGCTTGGCTTTAA 725

Qy 1068 ACATCCAGTTTATGG 1082

Db 726 ACATCCAGTTTGGG 740

RESULT 9  
BG501191 769 bp mRNA linear EST 27-MAR-2001  
LOCUS 60254624F1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4668562 5',  
DEFINITION mRNA sequence.  
ACCESSION BG501191  
VERSION BG501191.1 GI:13462708  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 769)  
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
plate: LLM41479 row: A column: 11  
High quality sequence stop: 758.



533 ATTGGCGTGAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGAGATTAATCATATC 592  
|||||  
244 ATTGGCGTGAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGAGATTAATCATATC 303  
|||||  
593 TTCAGAAATCCCAAGAGAGCTGACAGAGCCAAAGAGCTAGTGTGTAATATCAACAAAGGCT 652  
|||||  
304 TCCAGAAATCCCAAGAGAGCTGACAGAGCCAAAGAGCTAGTGTGTAATATCAATCAATAAAGGCT 363  
|||||  
653 GTGGCTATGGCTGTGACGCTCCATCATGTAGTGTACTGCTTTATGATGATATGACACCC 712  
|||||  
364 GTGGCTATGGCTGTGACGCTCCATCATGTAGTGTACTGCTTTATGATGATATGACACCC 423  
|||||  
713 AACGAACATCGCTGATGATCTCA-CAATGGGCGCTACGCTACGGGAGGAGGAATC 771  
|||||  
424 AACGAACATCGCTGATGATCTCA-CAATGGGCGCTACGCTACGGGAGGAGGAATC 483  
|||||  
772 GTGTTAGACCTGTAAAGAGAGCTGACAGAGATCTGGGAGCTCCACTGGACATTTGG 831  
|||||  
484 GTGTTAGACCTGTAAAGAGAGCTGACAGAGATCTGGGAGCTCCACTGGACATTTGG 543  
|||||  
832 TCAGGGAAGTAAAGAGCAAAATGTTCAAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 891  
|||||  
544 TCAGGGAAGTAAAGAGCAAAATGTTCAAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 603  
|||||  
892 CATCTCGCTCTCATATTTACCCCTGGGCTGTCAGAGAGCTGTCAGATGAGTGTGTA 951  
|||||  
604 CATCTCGCTCTCATATTTACCCCTGGGCTGTCAGAGAGCTGTCAGATGAGTGTGTA 662  
|||||  
952 CGAGTCCATGATGATCTGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 988  
|||||  
663 AGAGTCCATGATGATCTGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 699  
|||||

RESULT 11  
Bg771085 715 bp mRNA linear EST 15-MAY-2001  
LOCUS 602719363p1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4839758 5'  
DEFINITION mRNA sequence.

ACCESSION Bg771085  
VERSION Bg771085.1 GI:14081738  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 715)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgep@b-remail.nih.gov  
Tissue Procurement: DCTD/DIRP  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L16M1669 row: n column: 15  
High quality sequence stop: 665.  
Location/Qualifiers  
1..715

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="Caxon:3606"  
/clone="IMAGE:4839758"  
/clone\_lib="NIH\_MGC\_60"  
/tissue\_type="adenoecarcinoma"  
/lab\_host="DH10B (TI phase-resistant)"  
/note="Organ: prostate; Vector: pDNA-LIB (Clontech);  
Site\_1: SfiI (ggccgctcgccg); Site\_2: SfiI (ggccatctggcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGCGCATTTATGACC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGAGCGCGCGCGCGCGAGATG-dt(30)-BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
library.  
BASE COUNT 215 a 160 c 151 g 189 t  
ORIGIN

Query Match 32.2%; Score 557.2; DB 12; Length 715;  
Best Local Similarity 89.8%; Pred. No. 1,5e+128;  
Matches 643; Conservative 0; Mismatches 68; Indels 5; Gaps 4;

975 GTGGTGGGATCCAGCTTTGTCAAGTACTGATTTGCCCAACAACCTGGCTGGAAGA 1034  
|||||  
1 GTGGTGGGATCCAGCTTTGTCAAGTACTGATTTGCCCAACAACCTGGCTGGAAGA 60  
|||||  
1035 AATGAGAAGGCGCCAGAGAGCTAGGCTTCAACATCCAGTATTTGAGTCCATTTAG 1094  
|||||  
61 AATGAGAAGGCGCCAGAGAGCTAGGCTTCAACATCCAGTATTTGAGTCCATTTAG 120  
|||||  
1095 ACGCAGACAAAGAGGAGAGCGGAGAGCTTCCATTCATTTGAGGAAATACAGGTGCA 1154  
|||||  
121 ACGCAGACAAAGAGGAGAGCGGAGAGCTTCCATTCATTTGAGGAAATACAGGTGCA 180  
|||||  
1155 CGTTGAGAAGACTTTCAGCTTCTGCTGCGACAAATGCAAGTGGATTAAGGAGGTGTA 1214  
|||||  
181 TGTGGAAGAACTTTACGCTTCTGCGACAAATGCAAGTGGATTAAGGAGGTGTA 240  
|||||  
1215 TTGGGCGACAGATGACCTGCTTTGTTAAAGAGGAAACAAATACCCTGATTTA 1274  
|||||  
241 TTGGGCGACAGATGACCTGCTTTGTTAAAGAGGAAACAAATACCCTGATTTA 300  
|||||  
1275 ATTATTAAGATGATGATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1334  
|||||  
301 ATTATTAAGATGATGATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
|||||  
1335 TTGACCTCGGGGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATG 1394  
|||||  
361 TTGACCTCGGGGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATG 420  
|||||  
1395 TACTTTTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1454  
|||||  
421 TACTTTTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 480  
|||||  
1455 TGGCTCTGCGAAGCTTCCGTTTGTGATGATGATGATGATGATGATGATGATGATG 1514  
|||||  
481 TGGCTCTGCGAAGCTTCCGTTTGTGATGATGATGATGATGATGATGATGATGATG 540  
|||||  
1515 CAACCAATTTGCGATTTATCTGACCA-CTTCGAACTGAGAGAGAAATCCCATGGAAC 1573  
|||||  
541 CAATCAATTTGCGATTTATCTGACCAATCCCAATCCGAGATGGAATTTCCATGGAAC 600  
|||||  
1574 CTGGAGATTA-TTATTTGAGTGGCTGGAATCACTGGAGTGGCTATCTAAAGTGT -TA 1630  
|||||  
601 CTGGAGATTA-TTATTTGAGTGGCTGGAATCACTGGAGTGGCTATCTAAAGTGT 660  
|||||  
1631 ACAGAAACTGCGAAGAGAGGCGCTATATCCCTCTACAAAGTTGAGAGAGAGATA 1686  
|||||  
661 CCGGACAAATGGGAGAGGAGGCGCTATATCCCT-CTACAAAGTTGAGAGAGAGATA 715  
|||||

RESULT 12  
B0552954  
LOCUS B0552954 656 bp mRNA linear EST 20-JUN-2002  
DEFINITION H4019B10-5 N1A Mouse 7.4K cDNA clone set Mus musculus cDNA clone  
ACCESSION B0552954  
VERSION B0552954.1 GI:21453842  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM	TITLE	JOURNAL
Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 656)		
VanBuren, V., Piao, Y., Dudekula, D. B., Qian, Y., Carter, M. G., Marlin, P. R., Stegg, C. A., Bassey, U., Alba, K., Hamatani, T., Kargul, G. J., Luo, A. G. and Ko, M. S. H.		
Assembly, verification, and initial annotation of NIA 7.4k mouse cDNA clone set		
Unpublished (2002)		
Other_ESTRs: H4019B10-3		
Contact: Yong Qian		
Laboratory of Genetics		
National Institute on Aging/National Institutes of Health		
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6620, USA		
Email: cdna@igsun.grc.nia.nih.gov		
This clone set has been freely distributed to the community. Please visit <a href="http://19sun.grc.nia.nih.gov/CDNA/NIA_7_4k.html">http://19sun.grc.nia.nih.gov/CDNA/NIA_7_4k.html</a> for details.		
Plate: H4019 row: B column: 10		
Seq primer: -21M13 Reverse		
High quality sequence stop: 656		
POLYA-No.		

FEATURES	SOURCE
Location/Qualifiers	1. .656
/organism="Mus musculus"	
/strain="C57BL/6"	
/db_xref="niaesr:HA019B10-5"	
/db_xref="taxon:10090"	
/clone="HA019B10"	
/clone_1lb="NIA Mouse 7.4K cDNA Clone Set"	
/sex="mixed"	
/dev_stage="mixed"	
/lab_host="Dh10B"	
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."	
BASE COUNT	190 a 153 c 145 g 168 t
ORIGIN	

Query Match	32.2%	Score 555.8	DB 14	Length 656
Best Local Similarly	90.5%	Pred. No. 3.3e-128		
Matches 593	Conservative	0	Mismatches 62	Indels 0
				Gaps 0

[illegible]

Db	421	AAGGATATATCTGGCTACCTATATATCTTCTTGTTTAAAGAGGCAAGACAAAGTACTC	480
Qy	1266	CAGTTATGAATTTATTAAGTATGAATCTATCTCTTGGTCAGCTGGACATCAATATGATA	13235
Db	481	CAATTAATGAATTTTATAGTATAACTCTATTTCTTGGTCAGCTGGACATCAATCGGTA	540
Qy	1326	TACAGAAATTTCACTTGGGGGTGATCCTGGATATACACTTTCTCTCCAGGACGACTT	1385
Db	541	CACAGAAATTTCACTTGGGGGTGATCCTGGATATACACTTTCTCTACAGGCTGACCTT	600
Qy	1386	CGTAGTGTGACTTTTTCATCCGAGGCTGTAGTATAGTGTATTTGAAATCATGCA	1440
Db	601	TCTAGTGTGACTTTTTCATCCGAGGCTGTAGTATAGTGTATTTGAAATCATGCA	655

RESULT 13	
LOCUS	B0731540
DEFINITION	B0731540 900 bp mRNA linear EST 16-JUL-2002
ACCESSION	AGNCNCURT.8214839 NICHID XCC Embd4 Xenopus laevis cDNA clone
VERSION	IMAGE:J0742851.5', mRNA sequence.
KEYWORDS	B0731540 B0731540.1 GI:21870437
SOURCE	ESM
ORGANISM	African clawed frog. Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 900)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. Igor David  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.llnl.gov/bdnp/image/image.html](http://www-bio.llnl.gov/bdnp/image/image.html)  
plate: LLM10581 row: j column: 20  
High quality sequence stop: 607.

FEATURES	Location/Qualifiers
source	1. .900

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4724851"
/clone_1id="NICHD XGC Emb4"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/note="Organ: whole embryo; Vector: pCMV-Sport6; Site:1:
Not1; Site:2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

```

Query Match	32.0%	Score 552.4	DB 14	Length 900
Best Local Similarity	77.6%	Pred. No. 2.5e-127		
Matches 681	Conservative	0	Mismatches 196	Indels 1
			Gaps	1
Qy	492	GACGATCTATACACACACAGTGCAGTCAAAAGATGGGGCAGGTGTTGGCGTGAAAGAGAGC	551	
Db	1	GACAGATCTTCTACTCTCTCAGTCAGACAGATGGCCAGGCCACTGGCGTGAGAGAGAGGC	60	
Qy	552	CAAGAGATCTGACAGAGCGTGGTCCACGCGGAGATACATATCTTCCAGAAATCCCAAGAGCTG	611	
Db	61	CAAGAGATTTTGACAGATCTCTAGTACAAAGAAAGATCACTATCTACACGAACCCCAAGAGATTG	120	
Qy	612	CAGCAAGCCCAAGAAAGCTAGTGTGTATATATCAACAAGGCTGTGGCTATGCGCTGCACCT	671	



||||| 121 CAGCAAGCAAGAGCTGTTGTAATATCAACAAGGCTCGGGTATGGCTGCAGCT 180  
Db  
672 CCATCATATAGTACTCTCTTATATATGCAATATGGCACCCGCAACACTGCTTGA 731  
Qy  
181 CCACATATGTTGTAATCTTATATGTTATGTTGTTACGCAAAAGCCCTATCTTGA 240  
Db  
732 ATCTCACAATTTGGGCTGCTAGCTAGTGGGAGTGGAACTGTTAGACCTGTAAGTA 791  
Qy  
241 GTCCCAAGAGCTGAGAGTATGCCACTGTGTGATGGAAACCGTTTAAACCTGTACGCA 300  
Db  
792 GAGCTGACAGACAGATCTGGCACTCCAGTGCATTTGTCAGGTGGAAGTAAGACAA 851  
Qy  
301 CACCTGACAGAGTATGATCTGTTCTTACAGAGACACTGGGAGCTGAAGCAATGATA 360  
Db  
852 AAATGTTAGTGTGTTGAGTCCCTCATTTGTAGACAGTGTTCATCTCTCTCATATT 911  
Qy  
361 GAATGTTCAAGTGTGAGGATTAACCATTTGTACAGCCGTGATCCAGGCGCACTTACT 420  
Db  
912 ACCCTGCTGCTCCAGAGACCTTGACATGCACTGATGATGATGATGATGATGATG 971  
Qy  
421 ACCTTCTGGGCTGACAGAGAGCTAGCTGACCGCTTAACTGTGACAGGTGATCTGC 480  
Db  
972 AGTGTGTGTTATCCAGTTTGTCAAGTATGATGATGATGATGATGATGATGATG 1031  
Qy  
481 GGTGTGTGTTATCCAGTTTGTCAAGTATGATGATGATGATGATGATGATGATG 1091  
Db  
1092 TAGACGCAAGAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1151  
Qy  
601 CAGGAGAGAGGAGTAAAGTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1211  
Db  
1152 GCAGCTTGAAGAGAGCTTTCAGCTTCTGTCGAGAAATGCAAGTGAATAAAGAGGT 1271  
Qy  
661 GCATGTCGAGAGAACATTTCCACTGCTGTCGAGATGAGATGAGACAGAAACGAT 1331  
Db  
1212 GTATTTGGCCACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1368  
Qy  
721 TTTACTGCAACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1421  
Db  
1272 TGAATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1481  
Qy  
781 TGAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1541  
Db  
1332 AAATTT-CACCTTGGGAGTGTGATCTGATATACACTTT 1368  
Qy  
841 AAATTCCTCTCCGGGGGCTTATACGATATCCACTTT 878  
Db

RESULT 14  
Bg619839 653 bp mRNA linear EST 18-APR-2001  
LOCUS 602618014F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4731795 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg619839  
VERSION Bg619839.1 GI:13671210  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 653)  
AUTHORS NIH-MGC htlp://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: gsaabs-remail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.lnl.gov  
Plate: L1CM1590 row: 1 column: 04  
High quality sequence stop: 651.  
Location/Qualifiers  
1. 653  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4731795"  
/clone\_lib="NIH\_MGC\_79"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: placenta; Vector: pNR-LIB (Clontech);  
Site\_1: Still (ggcgccctggcc); Site\_2: Still (ggcgatagcc)  
); 5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCATTAATGCG-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGCGCGCGAGT-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
Library."

BASE COUNT 188 a 147 c 152 g 165 t 1 others  
ORIGIN

Query Match 31.1%; Score 536.8; DB 12; Length 653;  
Best Local Similarity 90.8%; Pred. No. 1.9e112;  
Matches 593; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

Qy 743 GCGGCTAGGCTCTGGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 802  
Db 1 GCGGCTAGGCTCTGGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60  
Qy 803 ACAGATCTGGGAGTCCACTGACATTTGGTCAAGTGAAGTAAGCAAAATGTTCAAG 862  
Db 61 ACAGATCTGGGAGTCCACTGACATTTGGTCAAGTGAAGTAAGCAAAATGTTCAAG 120  
Qy 863 TGTGTTGAGTCCCTCCATTTGATGACAGTGTTCATCTGCTCTCCATATTTACCCGCTG 922  
Db 121 TGTGTTGAGTCCCTCCATTTGATGACAGTGTTCATCTGCTCTCCATATTTACCCGCTG 180  
Qy 923 TCCCAAGAGCTTCCAGTGTGATGATGATGATGATGATGATGATGATGATGATG 982  
Db 181 TACCAAGAGCTTCCAGTGTGATGATGATGATGATGATGATGATGATGATGATG 240  
Qy 983 TATCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1042  
Db 241 TGTCTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
Qy 1043 AGGCCACCAAGAGCTTGGCTTCAAAATGATGATGATGATGATGATGATGATGATG 1102  
Db 301 AAGCCACCAAGAGCTTGGCTTCAAAATGATGATGATGATGATGATGATGATGATG 360  
Qy 1103 ACAAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1162  
Db 361 ACAAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
Qy 1163 AAGACTTTCAGCTTCTTGGCTGCAAGATGCAAGTGAATGAAAGAGGATGTTGGCA 1222  
Db 421 AACATTTTCAGCTTCTTGGCTGCAAGATGCAAGTGAATGAAAGAGGATGTTGGCA 480  
Qy 1223 CAGATGACCTGCTTGTGTTAAAGAGGCAAAAGCAAGTACCCAGTTATGCAAT -7ATT 1281  
Db 481 CAGATGACCTGCTTGTGTTAAAGAGGCAAAAGCAAGTACCCAGTTATGCAAT -7ATT 540  
Qy 1282 AGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1340  
Db 541 AGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Qy 1341 TCGGGGTGATCTGATATACACTTTCTCCAGGAGAGACTTCAAGTGT 1393  
Db 1393 TCGGGGTGATCTGATATACACTTTCTCCAGGAGAGACTTCAAGTGT 1393

